Museum Lighting, Colour Constancy and Melanopsin

Daniel Garside
ORCID: 0000-0002-4579-003X

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of University College London.

Civil, Environmental and Geomatic Engineering
University College London

December 17, 2019
I, Daniel Garside, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.
Abstract

Museums seek a pragmatic compromise between lighting which causes minimal damage to objects and lighting which allows maximal visitor enjoyment. One way to reduce damage is to choose illumination of a lower Correlated Colour Temperature (CCT), since this contains less short-wavelength radiation, which is generally more damaging.

Colour appearance models suggest that a change in CCT alone should not affect visual experience, but there is a long history of studies which aim to find a ‘preferred’ CCT for museum environments. These experiments have often had conflicting findings. One potential cause for this would be if the spectral sensitivity for adaptation differed from the spectral sensitivity for vision, as generally only the chromaticity (not the spectrum) is controlled in such experiments. Such a situation might arise if the recently discovered intrinsically photosensitive Retinal Ganglion Cells (ipRGCs), which express melanopsin, were involved in colour constancy.

Two psychophysical experiments were performed whereby observers were adapted to different spectra whilst performing an achromatic matching task. The first experiment used sixteen narrow-band light sources as adapting fields, and the second used two perceptually indistinguishable sources which differed in melanopic power. Neither experiment found evidence for a role of melanopsin in colour constancy, but it was noted that the model of colour constancy implicitly under examination was under-developed and did not produce clear predictions.

Regarding this, an exploratory computational study was performed, to explore whether, and in what way, a melanopsin signal might contribute to an observer’s ability to achieve colour constancy. It was found that a normalised melanopic signal...
provided a means by which colour constancy could be roughly achieved, without
the need for scene-level regularities which other algorithms rely upon.

Additionally, a novel method for performing colour constancy experiments
upon a tablet computer was developed, so that experiments could be run more
easily within real world environments.
Impact Statement

The applied goal of this research was to increase our understanding of colour constancy so as to advise museums on how to reduce damage to objects without degrading visitor experience. Asking whether intrinsically photosensitive Retinal Ganglion Cells (ipRGCs) play a role in colour constancy is additionally valuable to the vision science community, and to lighting engineering applications beyond the museum world.

In recent years there has been considerable take-up of Light-emitting diode (LED) lighting in museums, driven predominantly by energy-efficiency gains mandated at an institutional level. However, the spectrum of LEDs differs considerably from that of other lighting technologies, and so standard methods for limiting damage may be less effective than they have previously been. Code written for this thesis, and made available online, provides a simple means for computing damage predictions for specific light sources.

A series of interviews was performed with museum professionals to determine the current tools and methods which are in use by those making decisions regarding lighting. This information should allow lighting researchers and industry partners to better identify areas where improvement might be most beneficial, in terms of education, communication, and technological development.

There has recently been a great amount of interest in the lighting engineering community regarding the concept of human-centric lighting, which considers the impact of artificial lighting on the circadian rhythms of the inhabitants of a space. As such, the activation of ipRGCs is often increased (at least in the morning) compared to more traditional types of lighting. Knowing whether there is an additional visual
effect, which would need to be taken into consideration during design, would be of
great value to those designing such lighting systems.

Colour constancy algorithms have value outside of understanding the visual
systems of humans and other organisms. They are used extensively in imaging
devices, to maintain colour balance in most types of cameras, and to aid in ob-
ject recognition in computational vision applications. Providing more effective
algorithms to be used by such systems could substantially improve their overall
performance.

The work contained in this thesis, and produced as part of this doctoral pro-
gram, has the potential to be of impact to the fields of museum conservation,
lighting engineering, visual neuroscience, and digital imaging.
Acknowledgements

“The interlocking of friendship and science augurs well, if not just for the science, for the completeness of our lives, our all too limited time as scientists.”

Siegel [271]

This PhD program has allowed me many luxuries; an excuse to lounge in the reading room of the Wellcome Collection with an obscure book from one of the many libraries of UCL, for one, being provided with a light-tight basement (which for some reason no-one else wanted as a work-space) for another. The greatest luxury however, has been the connections with fellow researchers around the world. At conferences I have found camaraderie and unity in our shared fascination with the field of visual neuroscience.

“It is indeed wrong to think that the poetry of Nature’s moods in all their infinite variety is lost on one who observes them scientifically, for the habit of observation refines our sense of beauty and adds a brighter hue to the richly coloured background against which each separate fact is outlined.”

Minnaert [219]

Walking around a sunlit Boston with a friend recently, we agreed that whilst we had both learnt from books, seminars and experiments; a great deal of our learning had come from the years of wandering around the world, experiencing and thinking. Examining our visual systems through our own experiences has
helped us immeasurably. This is the mindset afforded by my time at University College London (UCL), under the encouragement of my supervisors.

I have the unusual privilege to have five supervisors, and each is deserving of my thanks. Stuart Robson (UCL), I thank for his calm and level-headed assistance in times of stress, and his confidence in me when it was needed. Kees Teunissen (Signify Research, née Philips Lighting Research) I thank for his patience, on the many occasions where my optimism morphed into tardiness. His scrupulous and careful feedback on my work often asked the hard questions which needed to be asked; generally the very same ones I was trying to avoid. Capucine Korenberg (The British Museum) I thank for her expertise and perseverance, particularly when advising me on the curious world of museum lighting, and arranging for me to be allowed to perform experiments at the British Museum. Katherine Curran (UCL) I thank for her valuable academic contributions as a collaborator, as well as her enduring support, kindness and positivity. Finally, I thank Lindsay MacDonald, who I met when he was a visiting professor at The University of Westminster (where I completed my undergraduate studies), delivering riveting lectures on colour science. After I finished my undergraduate degree, it was his encouragement that led to me to attend the AIC (Association Internationale de la Couleur) conference in Newcastle, where I confirmed my suspicion that the field of colour research was inhabited by interesting and unusual individuals and that I’d quite like to be part of it. It was at his encouragement again that I applied for the PhD position at UCL, and I have benefited from his support throughout, abusing his kindness in lending books from his library, and inheriting his questionably healthy habit of taking on many projects at once. I have had several enjoyable afternoons where we have spent hours in discussion, often drifting from the starting point into the mysterious corners of colour science.

My thanks go to the technical staff at UCL, in particularly the Chadwick Workshop technicians, who manufactured all manner of weird and wonderful things for me, the technicians at UCL PAMELA who supported me in using the space, and the staff of the UCL Grant Museum of Zoology for allowing me to use the
museum for initial experiments. I also wish to note my appreciation and thanks to the various pieces of open source software I used (for further details see Appendix A).

The research group of 3DImpact has changed greatly since I began. I have made many friends in this group, too many to mention. Thank you for your support and friendship.

My final thanks must go to my partner Alice, without whose patience and support this project would never have reached this point.
Contents

1 Introduction
   1.1 Context ................................................................. 23
   1.2 Research Question ................................................ 24
   1.3 Chapter Summaries ................................................. 24
   1.4 Associated Publications .......................................... 26
   1.5 Affiliations ......................................................... 28

2 Literature Review
   2.1 Colour Science ...................................................... 29
      2.1.1 Colorimetry .................................................... 30
      2.1.2 Colour Rendering ............................................. 42
      2.1.3 Chromatic Adaptation and Colour Constancy ............ 47
      2.1.4 Experimental Methods for Colour Constancy Research . 57
   2.2 Intrinsically Photosensitive Retinal Ganglion Cells (ipRGCs) . 65
      2.2.1 Synaptic Input ............................................... 68
      2.2.2 Projection ..................................................... 72
      2.2.3 The roles of ipRGCs beyond circadian entrainment . . . 72
      2.2.4 Value for colour constancy ................................ 76
   2.3 Museum Lighting .................................................... 79
      2.3.1 Damage Functions ............................................. 80
      2.3.2 Visitor Requirements of Museum Lighting ............... 86
      2.3.3 Museum Lighting Specification Guidance Documents . . 88
      2.3.4 CCT in Museums .............................................. 104
## Contents

2.3.5 Colour Rendering Indices in Museums .................................. 108
2.4 Mathematical Methods ......................................................... 110
2.4.1 K-means Clustering ............................................................ 110
2.4.2 Principal Component Analysis ............................................. 112
2.5 Interim Summary ................................................................. 114

3 Interviews with Museum Professionals 115

3.1 Introduction ................................................................. 115
3.2 Methodology ................................................................. 115
3.3 Summary of Responses ....................................................... 117
3.3.1 ’Good lighting’ ............................................................ 117
3.3.2 Roles and Tools ............................................................ 117
3.3.3 Quantifying Safety ...................................................... 118
3.3.4 Visual Testing .............................................................. 119
3.3.5 LED Usage ................................................................. 119
3.3.6 Spectral Tuning of Light Sources .................................... 121
3.3.7 Colour Rendering .......................................................... 122
3.4 Conclusion ................................................................. 123
3.5 Interim Summary ................................................................. 123

4 Large Sphere Experiment ......................................................... 125

4.1 Summary ................................................................. 125
4.2 Methodology ................................................................. 126
4.2.1 Hardware ................................................................. 126
4.2.2 Observer task ............................................................ 128
4.2.3 Stimulus specification .................................................. 130
4.2.4 Data Processing .......................................................... 131
4.3 Results ................................................................. 133
4.3.1 Summary ................................................................. 133
4.3.2 Variability over time/repeats .......................................... 135
4.3.3 Chromaticity-based analysis .......................................... 139
6.4.5 Environments ........................................ 216
6.5 Experiments ............................................. 224
  6.5.1 Experiment 1 ..................................... 224
  6.5.2 Experiment 2 ..................................... 224
  6.5.3 Experiment 3 ..................................... 224
6.6 Results ................................................ 227
  6.6.1 Experiment 1 ..................................... 227
  6.6.2 Experiment 2 ..................................... 230
  6.6.3 Experiment 3 ..................................... 236
6.7 Discussion ............................................. 239
  6.7.1 Environment-agnostic stimulus ..................... 239
  6.7.2 Meaningful data .................................. 239
  6.7.3 Naive observers .................................. 243
  6.7.4 Interacting with a large visual angle ............... 244
  6.7.5 Limitations of screen based experiments .......... 247
  6.7.6 Impact of ipRGCs on chromatic adaptation .......... 250
  6.7.7 Coding language .................................. 251
6.8 Conclusions ........................................... 251

7 Computational Study .................................... 253
  7.1 Summary .............................................. 253
  7.2 Introduction ......................................... 254
  7.3 Data sources ......................................... 255
    7.3.1 Illuminant and Surface Data Sources ............. 255
    7.3.2 Chosen Data sources ............................. 260
  7.4 Recovering the chromaticity of daylight .............. 261
    7.4.1 First-level signals ............................... 262
    7.4.2 Second-level signals ............................. 263
    7.4.3 A PCA interpretation ............................. 263
    7.4.4 Relationships as artefacts ....................... 267
  7.5 Transformation to an illuminant-independent space .... 269
## Contents

7.5.1 A melanopic signal as a third dimension .......................... 269  
7.5.2 Additive/Subtractive model ........................................ 272  
7.6 Assessment of colour constancy algorithms ....................... 277  
7.6.1 K-means clustering as an assessment tool ....................... 279  
7.6.2 A comparison of four algorithms .................................. 281  
7.6.3 A more naturalistic test case ...................................... 282  
7.7 Optimality of the melanopsin spectral sensitivity ............... 287  
7.8 Why is this possible? ................................................... 288  
7.9 Physiological plausibility .............................................. 290  
7.10 Implications for future research .................................... 292  
7.11 Conclusions ............................................................. 294  
7.12 Further work ............................................................ 296  

8 General Conclusions ....................................................... 299  
8.1 Summary of Conclusions ................................................ 299  
8.2 Contributions to Museum Lighting ................................... 300  
8.3 Contributions to Vision Science ....................................... 300  
8.4 Future work ............................................................... 301  

Appendices ................................................................. 304  
A Colophon ................................................................. 305  
B Ethics Application 9357/003 .............................................. 307  
C Ethics Application 9357/001 .............................................. 331  
D Ethics Application 9357/001 - Amendment .......................... 345  

References ................................................................. 387
## Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAM</td>
<td>Colour Appearance Model. 33, 278</td>
</tr>
<tr>
<td>CAT</td>
<td>chromatic adaptation transform. 52, 54, 57, 189, 278</td>
</tr>
<tr>
<td>CCI</td>
<td>Colour Constancy Index. 141–144</td>
</tr>
<tr>
<td>CCT</td>
<td>Correlated Colour Temperature. 5, 23, 24, 39–41, 44, 46, 86, 87, 89, 91, 94, 95, 97, 98, 102–106, 118, 121–123, 125, 263, 267, 299, 300</td>
</tr>
<tr>
<td>CMF</td>
<td>colour matching function. 31–33, 37</td>
</tr>
<tr>
<td>CQS</td>
<td>Colour Quality Scale. 100, 101</td>
</tr>
<tr>
<td>CRI</td>
<td>Colour Rendering Index. 23, 86, 87, 100, 109, 117, 118</td>
</tr>
<tr>
<td>DBUR</td>
<td>Difference Between Unannounced Repeats. 215, 216, 230, 236</td>
</tr>
<tr>
<td>DI</td>
<td>damage index. 23, 83, 85, 105, 106</td>
</tr>
<tr>
<td>GAI</td>
<td>Gamut Area Index. 100</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>IES</td>
<td>Illuminating Engineering Society. 45</td>
</tr>
<tr>
<td>ipRGC</td>
<td>intrinsically photosensitive Retinal Ganglion Cell. 5, 7, 24, 25, 65, 66, 68, 71, 72, 74, 76–78, 114, 125, 131, 148, 153, 155, 156, 161, 163, 173, 190, 197, 290, 294, 299, 300</td>
</tr>
<tr>
<td>LED</td>
<td>Light-emitting diode. 7, 44, 99, 100, 105, 117–121, 164, 170, 172, 210, 218, 225</td>
</tr>
<tr>
<td>MATLAB</td>
<td>Matrix Laboratory (MathWorks). 201, 214, 251, 280</td>
</tr>
<tr>
<td>NIF</td>
<td>non-image-forming. 65, 66, 72</td>
</tr>
<tr>
<td>PAMELA</td>
<td>Pedestrian Accessibility and Movement Environment Laboratory. 10, 209, 210, 216–219, 224–226, 239</td>
</tr>
<tr>
<td>PC1</td>
<td>the first principal component. 263, 266</td>
</tr>
<tr>
<td>PC2</td>
<td>the second principal component. 266, 275</td>
</tr>
<tr>
<td>PCA</td>
<td>principal component analysis. 112, 113, 263</td>
</tr>
<tr>
<td>PR650</td>
<td>Photo Research PR-650. 127, 166, 169, 186, 208</td>
</tr>
<tr>
<td>REM</td>
<td>Reflected Energy Matching. 97, 98</td>
</tr>
<tr>
<td>RGC</td>
<td>retinal ganglion cell. 65, 68, 78, 291</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>SAPS</td>
<td>Spatial Achromatic Point Setting. 191, 195, 196</td>
</tr>
<tr>
<td>SPD</td>
<td>spectral power distribution. 30–32, 49, 50, 56, 84, 86, 105, 106, 112, 120, 149, 170, 171, 216–218, 220, 294, 300</td>
</tr>
<tr>
<td>SRF</td>
<td>spectral reflectance function. 29–32, 47, 55, 56, 77, 113, 166, 168, 259, 260, 287, 289, 294, 296</td>
</tr>
<tr>
<td>SSF</td>
<td>spectral sensitivity function. 31, 33, 37, 144, 148, 149, 188, 287, 296</td>
</tr>
<tr>
<td>TCS</td>
<td>Test Colour Sample. 44, 45, 93</td>
</tr>
<tr>
<td>UCL</td>
<td>University College London. 9, 10, 126, 164, 197, 216, 219, 220</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet. 79, 81, 117</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Context

Light causes damage to objects in museums. Museums seek a pragmatic compromise between lighting which causes minimal damage to objects, and lighting which allows maximal visitor enjoyment. Generally this is achieved by following industry recommendations for maximum illuminance and Colour Rendering Index (CRI).

A complementary way to reduce damage, highlighted by the Commission Internationale d’Eclairage (CIE) in a 2004 publication [49], is to choose illumination of a lower Correlated Colour Temperature (CCT), because such illumination will contain most radiation energy in the longer wavelengths, which are generally less damaging to objects. To verify and extend the work of the CIE, a damage index (DI) was calculated for a range of commercially available lighting products (Section 2.3.4). This showed a strong correlation between predicted damage and CCT, with a factor of two between the low and high CCT sources. Practically, a sensitive object might be displayed for twice as long under a low CCT light source than under a high CCT light source, before passing a particular damage threshold.

A series of interviews with museum professionals showed that this technique is not currently being employed (Chapter 3). One reason is a belief that changing the CCT in an environment will affect the visual experience and the atmosphere in the room.

Most models of vision suggest that a change in CCT alone, within moderate
limits, should not affect visual experience; that in a space with a single type of lighting we should be able to adapt to any colour of illumination. This seems at most only partially true. It is true that we adapt reasonably well to the ambient illumination, both in terms of luminance and chromaticity, such that our perception of object colours relates primarily to object reflectance properties rather than the absolute intensities reaching our eyes. Generally, however, we do also seem to have an awareness of the properties of the illumination in a scene, and even a preference for some sources over others. Historically, experiments seeking to find an ideal preferred CCT have provided conflicting results.

One possible reason for these conflicting results might be if an additional retinal mechanism is involved in chromatic adaptation. The studies in this thesis have investigated whether a cell group called the intrinsically photosensitive Retinal Ganglion Cells (ipRGCs) might be involved in colour constancy. This cell group has traditionally been considered as having no output to visual pathways, and so colour constancy experiments haven’t controlled for ipRGC activation, but recent research has shown that ipRGCs do in fact have a limited input to visual pathways (see Section 2.2.3). There are a range of reasons which suggest that they may play a role in colour constancy specifically.

1.2 Research Question

The principal research question considered in this thesis is: Are ipRGCs involved in colour constancy?

If they are, this may explain why previous investigations into preferred CCT have produced conflicting results. If we can understand how they are involved, this may give us a new insight into how to vary the CCT of museum illumination in such a way that we can limit damage without degrading visitor experience.

1.3 Chapter Summaries

To better understand how museum professionals currently make lighting decisions (what metrics they use, what the guidance is for using these metrics, what the current challenges and limitations are) a series of interviews was performed
1.3. Chapter Summaries

with museum lighting professionals [123]. The responses to these interviews are summarised in Chapter 3.

To investigate the effect of different levels of ipRGC activation on an observer’s state of colour constancy, two lab-based psychophysical experiments were performed (Chapters 4 and 5). The first sought to examine the effect of different wavelengths of light upon chromatic adaptation. Within a Ganzfeld viewing environment, illuminated by one of 16 different wavelengths of near-monochromatic light, observers performed an achromatic setting task, controlling the chromaticity of a display visible in the central field through a 4° circular aperture with two handheld sliders.

In the second experiment the role of melanopsin in chromatic adaptation was more directly questioned. The same task was performed as in the first experiment, but the Ganzfeld was this time illuminated by one of two perceptually metameric lights with different melanopic illuminance levels. Neither of these experiments provided evidence for a simple or strong effect of ipRGC activation.

Concurrently, a method has been developed for performing colour constancy experiments outside of the lab environment, which can be completed quickly by naive observers in ‘real-world’ conditions (Chapter 6). The method uses a tablet computer, upon which an isoluminant plane through CIELUV is presented, successively varying in orientation and spatial offset. The observer is tasked with selecting, by touching with a finger, an achromatic point from within each stimulus. From the recorded selections an estimate of the observer’s state of chromatic adaptation is computed.

Following these experiments a different approach was adopted: rather than asking what was the effect of varying ipRGC activation upon experimental subjects, instead the question was posed whether an ipRGC-based signal could hypothetically be useful for colour constancy, considering what we know of daylight variability and natural surface reflectance properties (Chapter 7). A computational methodology was employed for this exploratory study. It was found that a simple transform can be made from raw signals to an illuminant independent space, with the use
of a melanopsin-based signal. Further, it was found that the spectral sensitivity of melanopsin is near-optimal for providing a signal for such a transformation, and that this can be done without using any scene level assumptions such as ‘grey-world’.

Chapter 8 provides a summary and discussion of results, and suggests routes for future work.

1.4 Associated Publications

Journal Papers


Conference Papers


1.4. Associated Publications


1.5 Affiliations

This project was supported by an EPSRC (Engineering and Physical Sciences Research Council) iCASE (Industrial Cooperative Awards in Science & Technology) studentship, with industry sponsors Signify Research (née Philips Lighting Research) and The British Museum.

Signify Research provided financial and technical support, with supervision from Kees Teunissen. The British Museum provided technical support and access to facilities, with supervision from Capucine Korenberg.

The project was supervised by Stuart Robson, Katherine Curran, Lindsay MacDonald, Kees Teunissen and Capucine Korenberg.

The thesis was examined by Anya Hurlbert (Newcastle University) and John Greenwood (University College London).
Chapter 2

Literature Review

This chapter shall provide summaries of the background information required to understand the content of this thesis, and also review the state of the art of the various subjects that this thesis covers.

2.1 Colour Science

Organisms on earth posses visual systems such that they can glean information about spatially remote objects through the sensing of light reflected from these objects towards the organism. They are able to do this because the sun emits electromagnetic radiation (which we call 'light' when it is within our visual sensitivity range), our atmosphere transmits many parts of this radiation (absorbing some wavelengths more so than others), and objects absorb some and reflect some of this radiation (again, some wavelengths more so than others), as shown in Figure 2.1.

Our perception of colour generally correlates with the way in which objects preferentially reflect some wavelengths over others (described by the spectral reflectance function (SRF)) which assists in the recognition of objects (that’s a banana) and the discrimination of distinct objects, often in a manner that is ecologically beneficial (that’s a ripe banana).
Figure 2.1: The spectral power distribution (SPD) of a single measurement of daylight (sunlight plus scattered blue-sky light, from Hernández-Andrés et al. [145]) and the light reflected from 2 different surfaces - a green banana and a yellow banana (data from personal correspondence with David Slaughter after Li et al. [192]), computed by multiplying the SPD by the measured spectral reflectance functions (SRFs) of the two surfaces. SPDs normalised such that the max of the daylight SPD is 1.

2.1.1 Colorimetry

The current recommended source for colorimetry is CIE document ‘CIE 015:2018’[56]. Whilst this is the authoritative reference, my personal opinion is that an understanding of CIE colorimetry is best gained from an understanding of its history, and for this I recommend Janos Schanda’s book ‘Colorimetry: Understanding the CIE System’ [265]. Another valuable, and perhaps more easily accessed, reference source is Stockman and Brainard [287].

I define ‘colorimetry’ as the study of the quantitative specification of colour.

\[\text{Equation from CIE 015:2018}\]

1 Though as Fairchild [97] notes, this document is ‘expensive and somewhat difficult to find’, and as such I have been using a draft of the now superseded ‘CIE 15.3:2004’[50] as my personal guide. All equations listed in this chapter come from this source.
As a subjective, internal and anthropocentric concept, in order to measure anything meaningful and comparable, we use a standard observer, or more precisely, one of a number of defined standard observers \[52\].

The classic standard observer was defined by the CIE in 1931, following experiments by Wright and Guild \[131, 323\]. Despite several more recently published standard observers, the 1931 observer is still much used, and I shall use it in the following example of how a basic colorimetric computation is performed.

An illuminant is defined by its spectral power distribution (SPD), a surface by its spectral reflectance function (SRF), and the sensitivity of a sensor (such as a photosensitive cell in the retina, or a pixel in a camera) by its colour matching function (CMF) (or in the case that biologically based measurements are used - the spectral sensitivity function (SSF)). Examples of these are shown in Figure 2.2.

The light reaching the eye for a given reflecting surface under a given illuminant (termed the ‘colour signal’) can be computed by multiplying the SPD by the SRF at each sampled interval.

\[
\phi(\lambda) = R(\lambda) \cdot S(\lambda) \tag{2.1}
\]

where \( R(\lambda) \) is the SRF, \( S(\lambda) \) is the SPD and \( \phi(\lambda) \) is the resulting colour signal. From this a set of values termed ‘tristimulus values’ may be computed. Tristimulus values are computed by multiplying the colour signal by the CMFs or SSFs, and following the principle of univariance (that each cone type cannot alone distinguish between different wavelengths) the tristimulus values can, to some extent, be thought of as containing the entirety of the chromatic information about a surface (for a human standard observer, excluding high level colour appearance phenomena...
Figure 2.2: An SPD (D65), a set of SRFs (from a Macbeth Colour Checker) and three CMFs (CIE 1931 observer).
predicted by Colour Appearance Models (CAMs), and properties such as gloss).

\[ X = k \sum_{\lambda} \phi(\lambda)\overline{x}(\lambda) \Delta \lambda \]  
\[ Y = k \sum_{\lambda} \phi(\lambda)\overline{y}(\lambda) \Delta \lambda \]  
\[ Z = k \sum_{\lambda} \phi(\lambda)\overline{z}(\lambda) \Delta \lambda \]  

where \( \overline{x}(\lambda) \) (said ‘x bar’), \( \overline{y}(\lambda) \) and \( \overline{z}(\lambda) \) are the CMFs of the 1931 observer (or are replaced by the CMFs/SSFs of the chosen observer), \( \phi(\lambda) \) is as defined above, \( k \) is a normalising factor, and \( \Delta \lambda \) is 1nm. It is traditional to set \( k \) such that \( Y = 100 \), though in the case of relative colorimetry (most cases) this is often not done since the following equation renders it superfluous.

From the tristimulus values \( (X, Y \text{ and } Z) \) chromaticity co-ordinates \( (x \text{ and } y) \) can be computed. These can then be plotted on a 1931 chromaticity space diagram, shown as Figure 2.3.

\[ x = \frac{X}{X + Y + Z} \]  
\[ y = \frac{Y}{X + Y + Z} \]  

2.1.1.1 Specific Colour Spaces

Over time various other colour spaces have been proposed and formally accepted by the CIE, each aiming to improve upon a prior space in one or more ways. One of the most frequently sought characteristics for a colour space is perceptual uniformity, whereby a set distance in one part of the space is comparable in terms of apparent colour difference to that same distance in another part of the space.

One such chromaticity space which shall be used extensively in this thesis is
the CIE 1976 UCS (uniform chromaticity scale) space (colloquially CIE u’v’), which is a relatively simple transformation of the 1931 chromaticity space. The conversion can be accomplished in a number of analogous fashions, one of which is shown below.

\[
\begin{align*}
    u' &= \frac{4x}{(-2x + 12y + 3)} \quad (2.4a) \\
    v' &= \frac{9y}{(-2x + 12y + 3)} \quad (2.4b)
\end{align*}
\]

where \( x \) and \( y \) are the 1931 chromaticity co-ordinates as defined in Equation 2.3. The data presented in Figure 2.3 are re-plotted in this new space in Figure 2.4.

A three-dimensional extension to the CIE 1976 UCS space is the CIE 1976 L’u’v’ colour space, often abbreviated to CIELUV, which aims to provide perceptual
uniformity in a space that includes both chromaticity and lightness (relative to a white object under the same illumination). It is defined as follows:

\[ L^* = 116f(Y/Y_n) - 16 \]  
\[ \text{where } f(Y/Y_n) = \begin{cases} (Y/Y_n)^{1/3} & \text{if } (Y/Y_n) > (24/116)^3 \\ (841/108)(Y/Y_n) + 16/116 & \text{if } (Y/Y_n) \leq (24/116)^3 \end{cases} \]  
\[ u^* = 13L^*(u' - u'_n) \]  
\[ v^* = 13L^*(v' - v'_n) \]  

where \( Y_n, u'_n \) and \( v'_n \) refer to the colour stimulus of a perfect reflector. The same data presented in Figures 2.3 and 2.4 is presented in CIELUV in Figure 2.5 (though of course the three-dimensional quality of the data is somewhat lost in this presentation).
Figure 2.5: A representation of CIELUV, showing the values for the Macbeth colour checker patches under D65 (as in Figure 2.3 and 2.4). Note that if viewed from above, the distribution of points would appear similar to as in Figure 2.4. If all of the points were of uniform $L^*$, then the distribution would be identical, only with different scaling and offsetting.

Introduced at the same time as CIELUV was a colourspace with a similar goal of perceptual uniformity: CIELAB. They are similar in many ways, and share $L^*$. CIELAB seems to be in slightly more common usage, but CIELUV is used in this thesis (particularly Chapter 6) due to the neat mathematical link to an associated chromaticity space (the CIE 1976 UCS space).

The two spaces suit subtly different purposes, requested by two different user groups identified at the time of the creation of these spaces: “lighting engineers interested in a space and formula that is a linear transformation of the CIE chromaticity diagram, and industrial colorists interested in a formula predicting perceived average color differences better than any existing one” [181]. CIELUV satisfied the first group, and CIELAB was designed to better suit the second, using a cube root function for the chromaticity values, mirroring $L^*$. CIELAB is defined
as follows:

\[ L^* = 116f(Y/Y_n) - 16 \]  \hspace{1cm} (2.6a)

where  \( f(Y/Y_n) = (Y/Y_n)^{1/3} \) if  \( (Y/Y_n) > (24/116)^3 \)  \hspace{1cm} (2.6b)

or  \( f(Y/Y_n) = (841/108)(Y/Y_n) + 16/116 \) if  \( (Y/Y_n) \leq (24/116)^3 \)  \hspace{1cm} (2.6c)

\[ a^* = 500[f(X/X_n) - f(Y/Y_n)] \]  \hspace{1cm} (2.6d)

\[ b^* = 200[f(Y/Y_n) - f(Z/Z_n)] \]  \hspace{1cm} (2.6e)

where  \( f(X/X_n) \) and  \( f(Z/Z_n) \) follow the same format as  \( f(Y/Y_n) \).

The final colour space presented here is the MacLeod-Boynton (MB) colour space. This space does not aspire to be perceptually uniform\(^2\). Instead, it aspires to provide a colour space which has a physiological underpinning. As such, the standard observer model consists of nominal SSFs (rather than CMFs), and the conversion from tristimulus values to chromaticity space attempts to mirror the way in which this actually occurs physiologically.

It was originally proposed by MacLeod and Boynton [208], but was recently revised and endorsed by the CIE, in documents ‘CIE 170-1:2006’ [51] and ‘CIE 170-2:2015’ [54]\(^3\). The revisions included the definition of a new observer (formally the ‘TC 1-36 Modified Colorimetric Observer’, colloquially referred to as the CIE 2006 observer), based on the work of Stockman et al. [289] and Stockman and Sharpe [288] (and referred to below as \( lms \)), and the introduction of a set of new chromaticities which can be clearly noted by comparing Figure 2.6 with previous figures. Where this set of chromaticities previously spanned the spaces fairly broadly, here all the chromaticities inhabit only a small portion of the chromaticity space.

\(^2\)This can be clearly noted by comparing Figure 2.6 with previous figures. Where this set of chromaticities previously spanned the spaces fairly broadly, here all the chromaticities inhabit only a small portion of the chromaticity space.

\(^3\)I understand that the final publication in the series (‘CIE 170-3:XXXX’) is in preparation.
normalising terms, such that it is now calculated as follows.

\[
L_{MB} = k_l \sum_{\lambda} \phi(\lambda)\overline{I}(\lambda) \Delta \lambda
\]  
(2.7a)

\[
M_{MB} = k_m \sum_{\lambda} \phi(\lambda)\overline{M}(\lambda) \Delta \lambda
\]  
(2.7b)

\[
S_{MB} = k_s \sum_{\lambda} \phi(\lambda)\overline{S}(\lambda) \Delta \lambda
\]  
(2.7c)

where \( k_l = 0.68990272 \), \( k_m = 0.34832189 \), \( k_s = 0.03715971 \). This set differs minimally but purposefully from Equation 2.2; in the choice of observer (\( \overline{LMS} \)), and in the use of different scaling factors for each tristimulus value. The different scaling values dictate the relative contributions of \( L_{MB} \) and \( M_{MB} \) to luminance, and constrict \( S_{MB} \) to a maximum value of 1 in the following equation (Equation 2.8).

MacLeod-Boynton (MB) chromaticity values are then calculated as follows:

\[
l_{MB} = \frac{L_{MB}}{L_{MB} + M_{MB}}
\]  
(2.8a)

\[
s_{MB} = \frac{S_{MB}}{L_{MB} + M_{MB}}
\]  
(2.8b)

Whereas the relationship between the axes of previous colour spaces was important, here the axes are nominally independent (representing different post-receptoral mechanisms) and so the scaling between the axes is arbitrary. As such, in vision science the axes of this diagram are often re-scaled to suit the task at hand (see for example Bosten et al. [20], Christiansen et al. [45], Danilova and Mollon [69]). The same data as presented previously is presented again in Figure 2.6 in a MB chromaticity space.
Figure 2.6: A MB chromaticity space diagram (as per CIE 170-2:2015 [54]) showing the values for the macbeth colour checker patches under D65, as in previous figures.

2.1.1.2 Correlated Colour Temperature

The Correlated Colour Temperature (CCT) of an illuminant is defined by CIE 15.3:2004 [50] as follows:

“The temperature of a Planckian radiator having the chromaticity nearest the chromaticity associated with the given spectral distribution on a diagram where the (CIE 1931 standard observer based) \(u'\), \(2/3v'\) coordinates of the Planckian locus and the test stimulus are depicted.”

Whilst by this definition a light source of any chromaticity could be assigned a CCT, as the distance between the Planckian locus and the test stimulus increases, the CCT becomes a less meaningful descriptor. CIE 15.3:2004 [50] defines chromaticity limits on the maximum distance between the Planckian locus and the test stimulus, beyond which CCT should not be used, which are described in Equation 2.9.
\[ \Delta C = \left[ \left( u_t' - u_P' \right)^2 + \frac{4}{9} \cdot \left( v_t' - v_P' \right)^2 \right]^{1/2} = 5 \cdot 10^{-2} \quad (2.9) \]

where \( u_t', v_t' \) refer to the test source and \( u_P', v_P' \) to the Planckian radiator.

Figure 2.7 shows the Planckian locus (the curve formed by a set of black body radiators of different temperatures) in CIE 1931 chromaticity space. It is worth noting that most real light sources (natural and artificial) fall upon the relatively straight section of the left-hand part of this line.

CCT is often used as the independent variable in experiments, with the implicit assumption that this can be considered a complete descriptor of a light source. This assumption only holds to the extent that the spectral form is similar between conditions (for example if only a single light source is used and well defined filtration is applied, which shifts the chromaticity of the light source along the Planckian locus), and that the only difference between conditions is a shift along the Planckian locus.

A shift in chromaticity perpendicular to the Planckian locus would result in no change to the CCT. To counter this issue, an additional parameter \( D_{uv} \) is defined [239] as per Equation 2.10, which provides a value for the distance from the Planckian locus.

\[ D_{uv} = \left[ \left( u' - u_0' \right)^2 + \left( \frac{2}{3} v' - \frac{2}{3} v_0' \right)^2 \right]^{1/2} \cdot \text{sgn} \left( v' - v_0' \right) \quad (2.10) \]

where \( \text{sgn}(z) = 1 \) for \( z \geq 0 \) and \( \text{sgn}(z) = -1 \) for \( z < 0 \).

Ohno [239] notes that the combination of CCT and \( D_{uv} \) is sufficient to describe the chromaticity of most light sources, and does so in a fashion which is slightly more intuitive than values of chromaticity.
Figure 2.7: The Planckian locus on a CIE 1931 chromaticity chart, highlighting CCTs of 1000K, 6500K and 10000K.
2.1.2 Colour Rendering

Colour rendering indices provide an indication of the colour appearance of objects under a test illumination. Colour fidelity indices, a subset of colour rendering indices\(^4\), are designed to describe how well a light source will produce a faithful appearance in terms of corresponding colorimetry to an appearance under CIE D-series illuminant (daylight proxy) or a black body radiator. Research in colour rendering is at a point of change and development, with the recognition that a single figure might never be enough to describe the complex, multidimensional, subjective and context dependent qualia of colour rendering [257]. Thus new indices, and modifications and addenda to long-established indices are being proposed [72, 158, 256, 273, 295].

If we were to take a bunch of balloons (or any objects, but balloons provide a simple and vivid image) of varied and multiple colours, from an area lit with daylight to an area lit with artificial light, and the colour appearance changes, for example the colours become more dull, we would generally be disappointed. In this case, we could say that we were dissatisfied by the colour rendering qualities of the artificial light source. If the colours remained the same we probably would think little of it, but if we were pressed for a comment, we might say that the colour appearance was satisfactory.

If the balloons stayed the same colour as they had been under daylight, we would be witnessing a high level of colour fidelity (fidelity meaning faithfulness or truthfulness). If fidelity aligns well with the priorities of the user, then high levels of fidelity are desirable. If the appearance of the balloons had changed in some other way than becoming more dull - for example the colours became more vivid or the hue of certain balloons changed drastically, we might not necessarily be disappointed, but this would still represent a situation where we were witnessing poor fidelity.

Generally when discussing colour fidelity, we are actually discussing fidelity

\(^4\)The most commonly used colour rendering index, CIE \(R_a\), is a colour fidelity index, which has led to the term ‘colour rendering index’ often being used when ‘colour fidelity index’ would be more correct.
2.1. Colour Science

to appearance under daylight or sometimes a Planckian radiator, aka a black body radiator. This is a fine comparison, but one which should be consciously considered; fidelity is not a measure of a light source per se, it is a comparison of that light source to a reference illuminant.

Fidelity has long been considered semi-analogous for colour rendering as a whole, but it is really only one element of colour rendering\(^5\). When considering how ‘good’ a light source is at rendering the colour of the objects which it illuminates, we might consider different measures alongside or in place of fidelity.

- **Colour Fidelity (to recap)** - A measure of comparison between how colours appear under daylight (or other reference) and how they appear under another light source.

- **Surface Discriminability** - The ability of a light source to illuminate objects such that subtle differences in colour are made most visible.

- **Observer Preference** - Subjective preferences for the colour of objects, potentially related to an increase in saturation, or an observer’s memory of colours, or specific colour effects (e.g. increased saturation of skin tones).

These areas overlap, and often exhibit correlation. For example, if a light source has properties which allow for high fidelity rendering, quite often that light source might also allow for good colour discrimination and also facilitate colours which might appear natural and/or pleasing. These are not reliable correlations however; it is quite possible to have a light source with very dissimilar properties to daylight that could score highly on discrimination and/or preference, or vice versa.

The index in widest use currently for quantifying colour rendering is CIE \(R_b\) (The General Colour Rendering Index). Until recently, this metric was used almost

---

\(^5\)Here I should point out that the CIE definition of colour rendering, from the publication CIE S 017/E:2011 ILV: International Lighting Vocabulary [53] reads: “Effect of an illuminant on the colour appearance of objects by conscious or subconscious comparison with their colour appearance under a reference illuminant.” You might notice the similarities to the definition I have ascribed to colour fidelity, and specifically not colour rendering. This is an unfortunate clash of nomenclature which I see as inappropriate usage on the part of CIE, and at some point I hope this official definition will be updated to match the modern vernacular.
exclusively to generate the figure listed as ‘CRI’ on the packaging for bulbs available commercially.

CIE $R_a$ is a colour fidelity metric, produced by computing the colour differences of 8 specified Test Colour Samples (TCSs), between the situation where illumination is provided by the test source and that where it is provided by a reference illuminant with the same CCT as the test source, and taking the arithmetic mean of these colour differences. The index is scaled such that the highest score is 100, for light sources which reproduce the TCS set under the reference illuminant exactly, with any light source which induces a colour shift receiving a value less than 100. There is no special significance for the zero value; it is quite possible (though uncommon) for a light source to receive a value of CIE $R_a$ below 0.

The most recent version of the standard specifying this metric is 'CIE 13.3-1995 Method of Measuring and Specifying Colour Rendering Properties of Light Sources' [48]. This document briefly overviews the history of, and necessity for, a colour rendering index and highlights areas of the procedure which vary compared to previous recommendations, and clearly details the recommended procedure. The process is summarised in the flow chart of Figure 2.8. For a worked example see Hunt and Pointer [152, p.388].

In response to a lack of confidence that CIE $R_a$ delivered meaningful values for Light-emitting diode (LED) illumination, a considerable amount of research and discussion has revolved around the subject of a colour rendering index to supersede CIE $R_a$. This is commonly complicated by the incorrect assumption that the values obtained from a fidelity index in some way correlate with preference. A more valid criticism of CIE $R_a$ is that it comprises deprecated methods (for example, use of the CIE 1960 UCS color space, now superseded) and data sets (a rather limited set of TCSs, drawn from measurements of Munsell papers) for computing intermediary stages within the process of calculation, and thus this theoretically limits the ‘correctness’ of any values produced by it.

Following many years where multiple CIE technical committees aimed to deliver updated colour rendering indices, but subsequently didn’t make any firm
recommendations, an Illuminating Engineering Society (IES) group was formed with the aim of examining the issue. The output of this IES group was the production of IES TM-30-15 [158], which follows the same philosophy as CIE \( R_a \) but with more modern colour science. It is still a contentious issue whether this updated index genuinely offers advantages over CIE \( R_a \).

Alternative propositions which offer fundamentally different approaches to the evaluation of colour rendering have been offered in recent years, including:

- indices based on production of memory colours [273].
- indices which consider the size of the gamut of a TCS set [257, 295].
- indices which are fundamentally fidelity indices but which penalise or reward specific traits known to be disliked/preferred [240].
Figure 2.8: The stages of calculation for the computation of the CIE General Colour Rendering Index. Notes: Reference illuminant: the same or nearly the same CCT. Below 5000K reference spectrum will be that of a black body radiator, and from 5000K 'one of a series of spectral power distributions of phases of daylight'.
2.1.3 Chromatic Adaptation and Colour Constancy

Shortly after the start of my PhD program I attended a symposium for PhD students in the field of lighting research (LumeNet). At this symposium Prof. Michael Pointer delivered a welcome talk, and one particular part of it stuck with me.

He described how at the start of his PhD studies, he had made copies of every paper which had been written on his subject (Chromatic Adaptation) which were available from the library, and written to any authors whose work he had been unable to access, requesting copies. At the end of his first year he had a binder which contained, as far as he was aware, everything ever written on the subject. He finished his story rather abruptly, along the lines of “since then, so many people have written on the subject that it would take much longer than you have within a PhD program to read it all. I don’t know what you all will do”. Not an encouraging thought!

With this in mind, I shall attempt only to cover the basics and a few select papers of particular relevance in the following section, and shall take this opportunity to note my debt to a number of comprehensive overview papers, book chapters and theses on this subject: Barnard [14], Brainard and Radonjić [26], Fairchild [96], Foster [113], Hurlbert [156], Maloney [212], Smithson [274].

2.1.3.1 The problem

At the start of this section I stated that “Our perception of colour generally correlates with the way in which objects preferentially reflect some wavelengths over others”. It transpires that this ability is non-trivial, because our visual systems do not have access to the SRFs of the objects which are seen, only the levels to which retinal cells are excited, which depends on the interaction of illuminant, surface and sensor.

Figure 2.9 demonstrates the problem, by reproducing Figure 2.3 with a minor adjustment; where the previous figure had the chromaticities of the set of surfaces under only a single illuminant, a large number of illuminants are used here. It can be seen that it is not uncommon for one surface to have a chromaticity under one illuminant that another surface has under a different illuminant. Thus, the relationship between surface and chromaticity appears to break down.

In order to provide a representation of colour where the perceived colour
remains constant across changes in illuminant (thus the term colour constancy), the human visual system must find a way to solve this problem.

2.1.3.2 Adaptation

‘Adaptation’ is the general mechanism by which a finite range of sensitivity can be shifted within absolute sensitivity bounds. The benefit of having an adaptive system, as opposed to a fixed system, is that the sensitivity of the system to small changes is maximised, whilst maintaining a broad overall sensitivity, at the expense of being able to sense over the entire range at a single time-point. A visual demonstration of this is shown in Figure 2.10 where it can be seen that at a single level of adaptation (a single line) the range of intensity over which responses are generated is relatively small, but is extended through adaptation.

In an environment such as the terrestrial environment, there is a great range in the level of illumination, but this range is rarely existent contiguously; levels
of illumination tend to be similar across a scene, and only change rather slowly over time. The notable exception, and thus where we notice the expense of having an adaptive visual system, comes when we enter or exit an environment where illumination is almost entirely excluded, such as a dark cave or below-decks of a boat.

Where the process of adaptation responds to overall illumination levels, it is referred to as light adaptation and dark adaptation. Where the process of adaptation responds to the wavelength composition of light reaching the eye it is referred to as chromatic adaptation.

In a natural environment, a change in the wavelength composition of radiation reaching the eye from a specific object might be caused by changing weather or time of day, or by the physical movement of the object from one space to another (where different lighting exists in the two different spaces). Figure 2.11 shows a set of measured daylight SPDs, normalised by luminance, showing that the wavelength composition varies quite considerably, though in a relatively systematic fashion.

2.1.3.3 Mechanistic vs. Inferential

In the literature there seems to be a long-running grapple between colour constancy and chromatic adaptation, with blurred boundaries and mixed definitions occurring
frequently. Brill and West [30] drive a wedge between the two concepts, but refute the traditional division that chromatic adaptation is the process and colour constancy is the ability, asserting that the two should be considered as separate processes, operating on different timescales and in different fashions.

Hurlbert [156] divides computational models of colour constancy into two classes: sensory and perceptual, which broadly follow the low-level / high-level divide. ‘Sensory’ models ‘could be achieved early in visual processing by adaptive scaling of the initial receptor responses’. ‘Perceptual’ models ‘would necessarily occur at later stages of visual processing’.

Hurlbert [personal communication] notes that understanding the differing perspectives of researchers from different fields allows one to understand how the mixed use of these terms in the literature may have arisen:

“The mixed use of the terms chromatic adaptation and colour constancy is largely explained by different usages in different fields. In
colour science, chromatic adaptation is often considered synonymous with colour constancy, as it is considered to be the only mechanism contributing to stabilising colour appearance under changes in illumination. In vision science, colour constancy is generally defined as a perceptual phenomenon, and distinction is made between different contributory mechanisms at different levels (which Hurlbert [156] attempts to define and onto which different proposed computational solutions are mapped).

Within vision science, the most common use of the term ‘chromatic adaptation’ now seems to be to refer to low level adaptation, with Brainard and Radonjić [26] using the term ‘mechanistic’ to refer to this type of transformation.

“Adaptation of this sort supports constancy to the extent that its overall effect is to stabilize the post-receptoral representation of the light reflected from objects across changes of illumination (as well as other contextual changes).”

The term ‘colour constancy’, by contrast, is used to refer to higher level inference-based processes, which occur at faster timescales than are possible by adaptation alone [258]. It seems likely that multiple processes work in tandem at multiple levels and timescales to achieve colour constancy [155].

2.1.3.4 Von Kries Adaptation

The conceptually simplest model of chromatic adaptation, and the forefather of many other models, is that referred to as Von Kries adaptation.

Johannes von Kries was one of the first to deeply consider chromatic adaptation, and appears to have understood it as a problem of sensory adaptation. In his formative series of papers, available in translation [311], he laid out how the study of chromatic adaptation could be divided into two broad and complementary aims.

The first referred to the systematic representation of the transforms in sensation caused by specific adaptations, such that “for any light mixture stimulating the readapted part [of the retina], another light mixture is specified that stimulates
the same sensation in a normally adapted part … The purpose of this study would be solved completely if a general rule could be obtained for the effects of all possible adaptations.” The search for ‘corresponding colours’ (pairs of colours which match under asymmetric adaptation states), and for suitable systems to predict the appearance of colours in all situations, fall within this category of exploration. Out of this aim grew the study of chromatic adaptation transforms (CATs), where the search for an algorithm which would mathematically predict the locations in colour space of corresponding colours has been the fundamental goal [57].

Von Kries suggested that the second categorical aim of chromatic adaptation research ought to be to discover ‘how the adaptation is produced by exposure to any particular color, continued over an extended period of time.’

Von Kries’ distinction carves a divide between the study of the mechanisms of chromatic adaptation, and the effects of chromatic adaptation. It would be unreasonable to think of these areas of study as unrelated, but it seems reasonable to recognise their separability.

Von Kries noted that ‘light mixtures that appear matched to the white-adapted eye always remain matched to the eye when it is adapted in any other manner.’ This is known to be only partly true; if referring to photopic vision solely (where rods are bleached and cones provide the basis for vision). The implication of this statement is that the spectral sensitivities of the cone cells do not change as a result of chromatic adaptation. If they were to change in some way, then distinct mixtures of light which match in one adaptive state might fail to match under a different state. Von Kries referred to this idea as the ‘persistence rule.’

The second rule which von Kries presents is referred to as the ‘theorem of proportionality’. This theorem claims that when two pairs of corresponding colours (colours A and B under illuminant 1, and colours C and D under illuminant 2) are additively combined (A with B, and C with D), the resulting colours will also be corresponding colours. It follows from the von Kries’ theorem of proportionality that increasing or decreasing the luminance of corresponding stimuli shouldn’t void their equality, though of course this statement relies even more heavily on the
caveat that this should only be assumed for photopic vision.

Von Kries goes on to suggest that it also follows that, if considered in combination with Grassmann’s law [129], the conversion of any arbitrary light by exposure to conditions causing a specific chromatic adaptation can be known if the nature of three other conversions are known. This relies on it being possible to consider any light as a linear combination of three others, and for the mechanism of chromatic adaptation to depend solely on the fatiguing of three independent systems.

Von Kries’ assertions may be mathematically described as:

\[
L_a = K_LL \tag{2.11a}
\]
\[
M_a = K_MM \tag{2.11b}
\]
\[
S_a = K_SS \tag{2.11c}
\]

where \( L, M \) and \( S \) represent the cone group responses; \( K_L, K_M \) and \( K_S \) represent distinct scalars and \( L_a, M_a \) and \( S_a \) represent the post adaptation cone group responses⁶. In terms of what might set these scalars, or gain values, von Kries suggested that:

“the organ of vision becomes less effective for that kind or for that part of its performance which is demanded from it for an extended period of time, whereas it becomes more effective for the activity which is, in a sense, opposed to that. This can be conceived in the sense that the individual components present in the organ of vision are completely independent of one another and each is fatigued or adapted exclusively according to its own function.”

Many interpretations have been made of this statement, and a corresponding number of nominations for methods by which to calculate values of \( K \) have been proposed. The most frequently considered methods are those referred to as ‘Grey World’ (GW) (where the reciprocal of the mean response across a scene is used) and

⁶This specific notation is taken from Fairchild [96, p. 183].
'Bright-is-White' (BiW) (where the reciprocal of the brightest point in the scene is used). Finlayson and Trezzi [108] showed that these two variants can be considered as part of a single framework (Minkowski norm), and that other members of this family of algorithms provide benefits over ‘Grey World’ (GW) and ‘Bright-is-White’ (BiW).

Von Kries’ thoughts have inspired many decades of research into CATs, which are a vital part of modern colour appearance models and other colour science computations, such as colour rendering index calculations. This relatively simple concept has remained relevant, as noted by Fairchild [96, p. 182] who reproduces this quote:

“*If some day it becomes possible to distinguish in an objective way the various effects of light by direct observation of the retina, people will perhaps recall with pitying smiles the efforts of previous decades which undertook to seek an understanding of the same phenomena by such lengthy detours.*”

*von Kries* [311]

followed by the comment:

“*Over eleven decades later, there is no one looking back at von Kries’ work with a "pitying smile". Rather, many are looking back at his work with astonishment at how well it has withstood the test of time.*”

*Fairchild* [96]

### 2.1.3.5 Retinex

Considered in the context on von Kries’ work, the Retinex model of Land et al. [186, 188–190, 215] might be described as ‘Von Kries adaptation with spatial considerations’ but I imagine such a statement would have rather riled the often bombastic sounding Land. It appears to have been Land’s understanding that the Retinex model was not a model of chromatic adaptation, but rather a model to usurp and do away with the very concept of chromatic adaptation (Land’s papers on the subject...
[186, 188–190] do not once use the term ‘chromatic adaptation’, nor is the work of von Kries ever explicitly referred to).

Land was fond of picking apart Newton’s statements regarding colour; particularly that the perception of colour was the result of an objects’ ‘excess and predominance in the [spectra of the] reflected light’ [236]. Land asserted that rather than work from the premise that the visual system was correcting an absolute record of the world, by normalising it to account for the ambient light source, one should instead consider visual input only as a relative record, where each element within a scene only takes on properties by comparison with the other elements of the scene.

He suggested for consideration the idea of the human visual system recording three separate lightness images, each representing the recording from a different band of receptors, where each element within each image is scaled against the brightest element in each specific image. Retinex theory also provides an interesting algorithm for discounting not only coloured illumination but illumination which varies across the visual field.

However, Barnard [14] points out that the various versions of the Retinex algorithm simplify to versions of the Von Kries adaptation, and Brainard and Wandell [27] found that ‘the algorithm is too sensitive to changes in the color of nearby objects to serve as an adequate model of human color constancy’. In addition, it has been found to be computationally difficult to implement, with no clear way in which it could be biologically implemented (though see Hurlbert [153]).

The work of Land et al. has spurred a great deal of progress in the world of digital imaging, where the term ‘computational colour constancy’ is used. The number of distinct algorithms increases at a rate which seems to always accelerate, and I shall not review them all here, though I shall point to valuable overview papers: Gijsenij et al. [127], Hordley and Finlayson [147], Hurlbert [156].

2.1.3.6 Linear models

If we extend the common interpretation of colour constancy (that colours should remain constant), to a stricter case where we actively try to recover the SRFs
of surfaces, simple Von Kries-type transformations no longer suffice. Maloney [210] and Yang and Maloney [326] discuss what they term the ‘RGB heuristic’. Simply put - this is the commonly held assumption that a set of cone catches under one illuminant will be related to a second set of cone catches under another by a transformation which can be fully described by the chromaticities of the illuminants. As Maloney [210] says ‘there is no mathematical reason to expect [this] to hold, even approximately’. This can be considered from the perspective of colour rendering; it is known that objects which are metamers under one illuminant might not be under another. This renders simple models of colour constancy ineffective. Foster et al. [114] quantified the frequency of metamerism in real scenes and found it to be ‘sufficiently large to affect visual inferences about material identity’.

For this stricter definition of colour constancy Maloney et al. [211, 212] propose a method which makes use of discrete linear models, whereby the set of real SRFs and real SPDs are approximated by a small number of fixed basis functions.

Maloney and Wandell [211] find that ‘with three classes of photoreceptors, we can exactly recover surface reflectances drawn from a fixed model of surface reflectance with at most two degrees of freedom’. Maloney [212] showed that this generalised such that with \( n \) classes of photoreceptors, SRFs could be reconstructed for surfaces drawn from models with \( n - 1 \) degrees of freedom, so long as there were not more than \( n \) degrees of freedom in the illumination model, and so long as the minimum complexity condition was met (there were enough surfaces visible).

Despite its mathematical elegance, quantitative analysis of this algorithm showed that its performance was poor [25, 107], which is not surprising considering that natural surface SRFs generally require more than two basis functions for accurate reconstruction.

There are many further ideas, theories, and concepts relating to colour constancy and chromatic adaptation which are beyond the scope of this chapter, but that the reader may find of interest: gamut mapping algorithms [105, 110–112], bayesian colour constancy [21, 22, 25, 106, 125] and specular-reflectance-based algorithms [156, 221, 223].
2.1.4 Experimental Methods for Colour Constancy Research

“How is colour constancy measured? With difficulty.”

Hurlbert [155]

The following section is divided into two following the split in research methods between those who could be described as coming from a ‘colour science’ perspective (where quantification and specification of colour are primary goals, and the development of CATs is a primary goal) and those coming from a ‘vision science’ perspective (where an understanding of the visual system is the primary goal). Further discussion of this split, and the introduction of a third group, comes in Section 7.6.

2.1.4.1 Within Colour Science

CIE 160:2004 [57] and Luo [198] provide a comprehensive overview of the different methods which have been used by the CAT group:

- Haploscopic matching
- Local Adaptation Matching
- Memory Matching
- Magnitude Estimation

**Haploscopic Matching** is the most common technique in this field, and the term refers to experiments which differentially adapt the two eyes and allow an observer to vary attributes of the stimulus presented to one eye such that it in some way matches the attributes of a fixed stimulus shown to another eye. Whilst this is in many ways unnatural, the benefits are that an experiment can be set up so that there is no time interference (the presentations are simultaneous, so memory effects are avoided) and that high precision of match is relatively easily achieved. An assumption is made that the adaptation of each eye is independent.
Local Adaptation Matching can be considered as a variation of haploscopic matching where instead of differing adaptational stimuli being presented to each eye, differing adaptational stimuli are presented to different parts of the same eye (spatially distinct areas of the same retina). The assumption here is that there is minimal intra-retinal interaction. MacAdam’s 1956 study [204] epitomises this technique. This experimental technique requires that observers minimise eye movement, in order to maintain spatially distinct adaptation.

Memory Matching has traditionally been performed by training observers to communicate colour sensation through the Munsell system notation [139, 185] and then asking observers adapted to different ambient lighting to describe set real objects using such notation. This technique is not much used due to many limitations and confounding factors. Luo [198] details the limitations of this experimental technique succinctly: ‘a substantial training period being required, complicated procedures for data analysis, lower precision than that of haploscopic technique, limited capacity for retaining information, and memory distortion.’

Magnitude Estimation appears to be similar to memory matching, in that observers are requested to verbally describe an object whilst in an ambient adapting field. The distinction is that the observers are requested to communicate their perceptions
2.1. Colour Science

using the perceptually meaningful attributes of hue, saturation and brightness, and as such results can be easily integrated into colour appearance models. Recent experiments [182, 199–202, 325] of this type collected data which was used to create CIECAM97s.

2.1.4.2 Within Vision Science

Methodologies similar to those described above, but with subtle yet important differences have been used by those whose primary goal is an understanding of the visual system. A valuable overview is provided by Foster [113], who lists four main methods:

1. Asymmetric color matching

2. Color naming and related methods

3. Achromatic adjustment

4. Discriminating illumination changes from reflectance changes

5. Illumination discrimination tasks

Asymmetric matching is in many ways analogous to the haploscopic matching described above, in that it describes an experimental set-up whereby one stimulus is compared with another, generally where each stimulus exists within a distinct adapting field and attributes of one stimulus can be either adjusted or responded to by an observer. The term asymmetric matching might be thought to be inclusive of a wider range of experimental set-ups, where haploscopic (Greek roots: haploieides, single and skopeo, to view) is necessarily concerned with each individual eye receiving distinct input. Asymmetric matching may refer to experiments where stimuli are viewed simultaneously, successively, or in an alternating fashion, binocularly or haploscopically.

Colour Naming is a technique employed with the aim being a more natural task than asymmetric matching, and removing some of the ‘instruction effects’ probed by Arend and Reeves [6]. Foster [113] argues that colour naming represents a task
apt to measure colour constancy more directly, as opposed to the ‘relational colour constancy’ often studied in asymmetric matching experiments, since it concerns identification rather than equivalence. One clear benefit seems to be that the observer needn’t be aware of the equivalence; it is expected that in such experiments there will be only one stimulus, perhaps with a temporally variant adapting field. An observer is simply asked to name colours, and this should theoretically result in a measure of adaptational colour constancy as opposed to inferential colour constancy, so long as the stimuli is suitably abstract. Colour names may be of a fixed set, or an observer may be given free choice. Analysis of results can employ a naming centroid based approach or a boundary focused approach. Speigle and Brainard [277] proposed a novel approach combining magnitude estimation and colour naming with the aim to improve precision of response.

**Achromatic Adjustment** experiments ask an observer to set a stimulus to a neutral achromatic hue, on the assumption that the internal grey point of an observer shifts in response to different adapting fields. These adjustments are generally easy for an observer to make, but the extrapolation of the results makes various assumptions about conceptual colour space and the nature of achromacy. Experiments are easily confounded by complex or real stimuli where there exists a close-to-neutral object in the scene which could consciously or unconsciously be used as a reference.

**Discriminating illumination changes from reflectance changes** provides a key way to examine colour constancy in an operational manner. Following the assumption that chromatic adaptation allows an observer to discount the illumination in some manner, an experimental set up where observers are requested to distinguish between an illumination change and a reflectance change represents a situation which very closely mirrors the natural process of colour constancy. This experimental technique is well placed to examine whether or not colour constancy in this form is active and efficient, but it provides little way of probing the underlying mechanisms of colour constancy.

**Illumination discrimination tasks** comprise a scene which is unvarying in terms of content and surface reflectances, but varying in terms of illumination. In the
2.1. Colour Science

A general set-up (such as that described by Pearce et al. [247]), a target illumination is compared to two test illuminations, where one of the test illuminations is identical to the target illumination and the other differs in chromaticity. The observer performs a two-alternative forced choice task whereby they attempt to identify the identical illumination, and from this estimates of the observers discrimination thresholds for illuminations of differing chromatic directions away from the test illumination can be estimated. Aston et al. [8] describes the interpretation of data from such experiments as offering two types of insight. Firstly, where no difference can be seen by observers this can be thought of as perfect colour constancy (assuming that the chromaticity differences are above surface discriminability thresholds) and comparisons across different chromatic directions can be made. Secondly, an indication as to the accuracy with which illumination colour is encoded can be gleaned. Concerns regarding this methodology were recently raised by Weiss et al. [318] who “did not find any relationship between achromatic adjustments and illumination discrimination thresholds” which “casts doubt on the idea that illumination discrimination directly translates into colour constancy”.

2.1.4.3 Achromatic Adjustments under Different Illuminations

A number of researchers have performed experiments similar in nature to those reported in Chapters 4 and 5, and these are summarised below. These references were added following suggestions by the examining committee for this thesis.

**Werner and Walraven [320]** performed an achromatic adjustment experiment whereby chromatic annuli (60°-90°) were presented as flashed stimuli (3 seconds on, 3 seconds off) on top of a background pedestal of varying chromaticity, and the observers task was to modify the chromaticity of the annuli such that they appeared achromatic.

The authors found that the luminance contrast between the background and the annuli had an effect; the higher the contrast between the two, the less saturated the annuli needed to be to appear achromatic. The authors note that “[t]his is not what should be expected on the basis of a simple von Kries transformation. The

---

7More succinctly reported in Walraven and Werner [315].
latter should yield, for a given background, hue shifts that are the same for dim (low contrast) or bright (high contrast) test stimuli” [315].

The authors also found an effect of the background luminance; higher luminances required more saturated annuli to appear achromatic.

The authors found that by considering only the chromaticity of the additional annulus (rather than the sum of annulus and background) they could account for the contrast effect (the background luminance effect remaining). Following this type of computation, they could account for the adaptive shifts with receptor-specific gain controls.

Brainard [24] used a method of achromatic adjustment in a slightly more naturalistic condition; a room was built where the ambient lighting could be carefully controlled, and where the chromaticity of a target patch could be controlled by an observer, through use of a projection colorimeter. Munsell papers and ‘theater stage lamps’ were used, and so this set-up was still considerably non-naturalistic compared to a real world environment.

The key findings from this series of experiments were that generally the constancy indices were very high (observers exhibited strong and effective colour constancy), and that the immediate surround of a target patch had very little effect on the achromatic point selected. This second result “differs [s] from a number of studies in the literature in which local contrast decisively determined appearance”. The reason for this difference was proposed as the greatly increased “richness of the stimulus conditions.”

The authors also found no discernible effect of the chromaticity of the illumination (in terms of one inducing adaptation to a greater extent than another). This “contradicts the idea, suggested by a number of theorists, that the visual system could take advantage of the fact that most natural illumination variation is along the daylight locus”.

In contrast to the findings of Werner and Walraven [320], as noted by Aston and Hurlbert [7], Brainard [24] “found that achromatic settings made at fixed luminance levels lie along a straight line in a three-dimensional cone space, concluding from
this that changing [target] luminance did not affect the chromaticity setting of an
observer’s achromatic point” [7].

Kuriki [183] starts by referring to the above finding of Brainard [24], but goes
on to note that ‘many studies have shown that the color appearance of a light
of fixed chromaticity varies with its intensity’, referring directly to the results of
Werner and Walraven [320] amongst others, and offering the Helson-Judd effect
(wherby ‘lighter surfaces tend to take on the color of the illuminant’) as a concrete
example whereby appearance can be seen to change as a direct result of changes
in the intensity of surface reflectance.

The author suggests two possible reasons for this discrepancy: the difference
between experiments dealing with increments and decrements, and the use of
display technologies vs. ‘real’ scenes.

The results from an achromatic setting experiment are described, where a
real-world (room-based) environment was used with a screen-based target (though
the target is said to have appeared as a piece of paper on the wall). A fitting to the
data from this experiment suggests that “chromatic adaptation is not describable
as a change in sensitivity parameters alone for an otherwise linear system. The im-
plied nonlinearity takes the form of an adaptation-dependent change in exponent”,
broadly agreeing with the findings of Walraven and Werner [315]. The authors
note however that the predictions of the model developed by Walraven and Werner
[315] do not “agree well with the present data”.

The author also further compares the results to those of Brainard [24]:

“ The results of Brainard’s study on achromatic loci (Brainard [24])
agree with ours in some points, but he found no pronounced intensity
dependence of the entries of the weighting coefficients for the cone
excitations. A possible cause of the difference is that the present study
contained more higher-lightness conditions than his. ”

Weiss et al. [318] report two experiments, one using a colour naming method
and one using an achromatic adjustment method, both performed using screen-
based stimuli for 40 different chromatic illuminants.
They test a large number of theories of colour constancy (referring to each as a 'determinant'), and found limited support for a role of colour naming and a strong blue bias. Additionally, “[t]here was some evidence that other factors, such as metamer mismatching, relational colour constancy, and colour variegation, play a role in colour constancy, but in a rather complicated way. In any case, the blue bias and the consistency of the illumination categories explained most of the variance of the achromatic adjustments.”
2.2 Intrinsically Photosensitive Retinal Ganglion Cells (ipRGCs)

Recent reviews by Spitschan [278], Do [81], Graham and Wong [128], and Lucas [195] provide authoritative overviews of this rapidly progressing research area. This section shall focus on the subgroup of ipRGCs called ‘M1’ ipRGCs, since these are the most populous and best understood; for details regarding the other subgroups of ipRGCs see Ecker et al. [90].

Retinal rod and cone cells (of three types, l/m/s) are well established as the primary receptors for human vision, and their connections and properties are relatively well understood. Two modes of vision originate from the retina, one of which is associated with image formation and the other which is considered to be non-image-forming (NIF), and which influences systems such as circadian rhythm entrainment, pupillary reflex and melatonin release. It was originally thought that rods and cones were the sole inputs to both of these modes of vision [134].

Retinal ganglion cells (RGCs) combine signals from groups of cones and rods and relay these signals via the optic nerve to the lateral geniculate nucleus, which in turn processes and relays them further to the cortex for additional processing, allowing for classical vision of objects, movement and colour. IpRGCs are a sub-class of RGCs, which in addition to combining and relaying signals exhibit some intrinsic photosensitivity of their own.

This intrinsic photosensitivity was only confirmed recently relative to our knowledge of other retinal cell types [254], following a search for a retinal cell type or combination of cell types which would fit the spectral sensitivity properties found to influence entrainment of the circadian rhythm in humans and other animals [28, 29], which was dissimilar to all of the spectral sensitivities of the cell classes known at the time.

Additionally, it was found that animals and humans with no functioning rods or cones were still able to have a correctly functioning circadian system [118, 327], further suggesting that that the circadian rhythm was influenced by a novel receptor
with a distinct photoreceptor. It is now believed to be these cells which provide the primary input to the NIF pathway.

IpRGCs were found to express a photopigment fitting such attributes, and it was given the name ‘melanopsin’. The spectral sensitivity of melanopsin peaks at around 480nm [12, 68, 133, 249, 254] which places it between the s-cone (cyanolabe photopsin) and rod cell (rhodopic rhodopsin) spectral sensitivities, see Figure 2.13. In humans, pre-receptoral filtering leads to a functional peak sensitivity of closer to 490nm [55].

Exogenously expressed mouse melanopsin has been shown to be tristable [93, 213], that is: existing in one of three possible states. A photon interaction converts melanopsin in one of these states to another, with two of the states being electrically silent (not providing a signal) and one being signal-producing. Notably, these different states have slightly different spectral sensitivities, and thus exposure to specific wavelengths biases the population distribution in different ways, as shown in Figure 2.14.

Human melanopsin appears to be bistable (of two states), and there is
"Mouse melanopsin is understood to have three states (R, M, and E). The peak spectral sensitivities of R and E are determined from the electrophysiological responses of M1s. Spectrophotometric measurements of purified melanopsin yielded similar values (467 and 446 nm, respectively) and gave information for M (476 nm; [Matsuyama et al. [213]]). Bottom, the distribution of melanopsin states as a function of wavelength, estimated from a model based on values from purified melanopsin."
uncertainty regarding whether this bistability has physiological consequences [55, 93, 175, 214, 228, 229, 260, 314]. It has been shown in other organisms that colour opponency from a single opsin is possible [314].

IpRGCs form a sparse mesh across the retina, each covering roughly 10° of visual angle [90]; discounting input from other cell types, they operate at a much lower resolution than as would be required for spatial vision of the type we are accustomed to.

They also operate much more slowly than other cell types, taking several seconds to respond, but are able to sustain a response in contrast to other retinal cell types which are able to respond quickly but only for short periods (See Figure 2.15 and Do [81, p.210] for a summary).

IpRGCs vary greatly in the range of light intensities that they are sensitive to, are unimodal (stop responding above a certain threshold), and as a population their sensitivity spans a large range of light levels [81]. It has been proposed that these properties ipRGCs would allow an observer to efficiently sense a level of absolute irradiance [35, 218] (restricted to the wavelengths which ipRGCs and their inputs are sensitive to).

2.2.1 Synaptic Input

In addition to their intrinsic photosensitivity, ipRGCs exhibit extrinsic photosensitivity, taking inputs from rod and cone pathways in a similar fashion to regular RGCs. Synaptic input is provided by amacrine and bipolar cells (See Belenky et al. [17] and Figures 2.16 and 2.17).
Schematic of the relevant retinal circuitry in humans. Non-image-forming responses originate in the retina and have been attributed to a particular class of retinal ganglion cell (ipRGC). ipRGCs are directly photosensitive owing to expression of melanopsin, which allows them to respond to light even when isolated from the rest of the retina. In situ they are connected to the outer retinal rod and cone photoreceptors via the conventional retinal circuitry. The details of their intraretinal connections are not completely understood and probably vary between different subtypes. Shown here are major connections with on cone bipolar cells (on CBCs) connecting them to cone and, via amacrine cells (AII) and rod bipolar cells (RBC), rod photoreceptors. As a consequence, the firing pattern of ipRGCs can be influenced by both intrinsic melanopsin photoreception and extrinsic signals originating in rods and each of the spectrally distinct cone classes (shown in red, green, and blue).
Figure 2.17: Reproduced from Do [81] (Fig 1A in original source). ‘A highly simplified schematic of the retina in cross-section, oriented with the inner aspect (nearer the center of the eye) down. The outer photoreceptors (i.e., rods/cones) drive bipolar cells (BCs). In the inner plexiform layer (IPL), BCs synapse with retinal ganglion cells (RGCs). Left: ON circuitry. Rods (top) drive rod BCs, whose signals pass through amacrine cells (ACs) to cone BCs. In the inner IPL, ON cone BCs convey signals to ON RGCs. ON RGCs show greater depolarization when light intensity increases. Center, OFF circuitry. In the outer IPL, OFF cone BCs provide synaptic input to OFF RGCs. OFF RGCs show greater depolarization when light intensity decreases. Right: a sample of circuits for outer- and inner-stratifying ipRGCs, which are both ON. ON cone BCs make ectopic synapses with the former and conventional synapses with the latter. Rod pathways also drive ipRGCs. IpRGCs make chemical and electrical synapses with ACs (not shown).”

Inputs evoke ON responses, despite originating from both the ON and OFF layers of the inner plexiform layer (see Figure 2.17). Graham and Wong [128] describe the processes which allow this to occur:

“ON bipolar cells use two unconventional strategies to release glutamate onto the dendrites of SCN-projecting mouse ipRGCs in the “OFF” sublamina. Some ON bipolar cells’ axons extend lateral protrusions that contain synaptic vesicles […], whereas others possess en passant (in passing) synaptic vesicles within their axonal shafts [Dumitrescu et al.
2.2. Intrinsically Photosensitive Retinal Ganglion Cells (ipRGCs) [84]. Distally stratifying ipRGCs are also present in rabbit, marmoset and macaque retinas, and they likewise receive unconventional ON bipolar input in the “OFF” sublamina [...] [Hoshi et al. [149], Grüner et al. [130]].

Signals from rods and cones retain their traditional time courses; ipRGCs are not inherently sluggish, rather the melanopic inputs to ipRGCs are. This can be seen in Figure 2.18, where standard outputs of an ipRGC are shown on the left, and outputs with rod/cone driven synaptic inputs blocked are shown on the right. The response is shown to be relatively instantaneous for the cell with intact inputs, but lagging by several seconds for the cell relying on melanopic activation alone.

**Figure 2.18:** Reproduced from Wong et al. [322] (Fig 4 in original source). “Multi-electrode array (MEA) recordings of synaptically mediated light responses in ipRGCs. Extracellular recordings comparing spike responses of an ipRGC to light of various intensities with rod/cone-driven synaptic inputs left intact (left) or blocked (right). Log stimulus attenuation is indicated to the left. Note that all responses to weaker stimuli (-2 log attenuation and dimmer) and short-latency responses to brighter ones (-1 to +0.65 log I) were dependent on synaptic transmission, presumably because they reflect rod and/or cone influence on the recorded cell. Note also that intensities sufficient to recruit the intrinsic response (-0 and +0.65 log I; right) evoke responses with substantial poststimulus persistence, a well-established feature of melanopsin-dependent light responses.”
There is also evidence for ipRGCs having intraretinal retrograde synaptic output (Zhang et al. [331, 332], summarised by Graham and Wong [128]), via a subpopulation of dopaminergic amacrine cells. This type of signalling could provide a feedback loop which modifies signals before they have left the retina.

2.2.2 Projection

ipRGCs have been shown to innervate ‘dozens of brain areas’ [81], with M1s principally innervating NIF areas, but with some activation of areas traditionally thought of as image-forming. There appear to be meaningful differences between the different sub-types of ipRGC in this respect, with different sub-types (denoted M1-6, distinguished by their retinal morphology) showing distinct activation pathways.

A summary of these projections is shown in Figure 2.19. From this figure it can be seen that there are still many potential projections which have not been investigated (“Each blue dot indicates the approximate density of innervation by its size, a white dot indicates undetectable innervation, and lack of a dot indicates an absence of information.”). However, it can clearly be seen that outputs are extensive and varied.

2.2.3 The roles of ipRGCs beyond circadian entrainment

In recent years there has been a number of publications examining the role of melanopsin outside of NIF vision, challenging some of the assumptions about the roles of the signals originating from ipRGCs.

A number of studies have found that the signals from ipRGCs are capable of encoding spatial structure [3, 4, 90, 227, 334] and others have probed the influence upon brightness perception [36, 328].

Additionally, several researchers have investigated whether ipRGCs may play a role in chromatic vision [38, 148, 285, 309, 310, 329], using the silent substitution paradigm [95, 280]. These studies are summarised below.

Spitschan et al. [285] found an fMRI response in primary visual cortex for each

---

8 For commentary see Spitschan and Aguirre [279] and Sonoda and Schmidt [276].
9 Though see Kamar et al. [166].
2.2. Intrinsically Photosensitive Retinal Ganglion Cells (ipRGCs)

Figure 2.19: Reproduced from Do [81] (Fig 3 in original source). “Major Brain Targets of Mouse IpRGCs. A sample of ipRGC brain targets is depicted in a quasi-sagittal schematic of the mouse brain. Below is a plot of innervation densities across ipRGC types, drawn after Berson and colleagues (Quattrocchi et al. [255]) and incorporating additional information (Ecker et al. [90]; Hattar et al. [137]; Huang et al. [151]; Morin and Studholme [224]; Zhao et al. [334]). Each blue dot indicates the approximate density of innervation by its size, a white dot indicates undetectable innervation, and lack of a dot indicates an absence of information. M5s and M6s are pooled because their projections were examined together for technical reasons. AH, anterior hypothalamus; BST, bed nucleus of the stria terminalis; dLGN, dorsal lateral geniculate nucleus; IGL, intergeniculate leaflet; LH, lateral hypothalamus; MA, medial amygdala; OPN, olivary pretectal nucleus (with shell, s, and core, c, regions); PA, preoptic area, which includes the VLPO (ventrolateral preoptic area); PAG, periaqueductal gray; PHb, perihabenular zone; pSON, perisupraoptic nucleus; SC, superior colliculus; SCN, suprachiasmatic nucleus; sPa, subparaventricular zone; and vLGN, ventral lateral geniculate nucleus.”

of four participants, in contradiction to an earlier study by the same group [283]. Participants reported a visual percept which was “unpleasant, blurry, minimal brightening that quickly faded”. There was also some evidence of a chromatic percept: “Many of the subjects described the melanopsin stimulus pulse as being colored. This was typically a yellow-orange appearance, although three subjects reported a greenish percept”.

Cao et al. [38] found that “changing melanopsin activation levels shifts the
equilibrium point in the chromatic pathways”, though curiously the effect was only present for the L/(L+M) pathway. The key figure from this study is reproduced in Figure 2.20.

The authors conclude that melanopsin activation affects the parvocellular pathway to the extent that ipRGC activation could be thought of as “additive to the M-cone signal opposing the L-cone signal in the PC pathway [i.e., L - (M + I)] (where “I” for melanopsin activation in ipRGCs) to signal greenness and/or blueness”.

The authors note that this corresponds to an earlier result from one of the same authors [16] where such a contribution set was proposed. However, the earlier result includes rod contributions, and the specific pathway which they must be referring to is only the 5th component accounting for < 0.01% of the variance in the signals under examination. Meanwhile, no evidence is found for the 2nd component from that same analysis (labelled as konioncellular, representing 1.56% of variance), which they also proposed would have a considerable melanopic contribution. They neglect to mention this in the later paper.

\[ Z \text{e}l\text{e} \text{ et al. [329]} \text{ found evidence that “putative melanopsin-mediated image-} \]
\[ \text{forming vision corresponds to an opponent S-OFF L+M-ON response property,} \]
2.2. **Intrinsically Photosensitive Retinal Ganglion Cells (ipRGCs)**

with an average temporal resolution up to approximately 5 Hz, and >10x higher thresholds than red-green colour vision”. The key figure from this study is reproduced in Figure 2.21. This figure shows the perceptual matches to melanopic contrasts in terms of equivalent cone contrasts, for four observers, under three different conditions. For each observer, and for each condition, it can be seen that a melanopic contrast can be matched by an L+M increment and a S/(L+M) decrement.

Figure 2.21: Reproduced from Zele et al. [329]. “Melanopsin photoreception is analogous to an increment in cone luminance [L + M] and a decrement in S-cone excitation [S/(L + M)] with white (a), yellowish-pink (b) and orange adapting stimulus fields (c). […] Data in each panel are for four participants (mean ± SEM) measured at 2000 photopic Td.”

This appears to be a perfect contradiction to the results of Cao et al. [38]: Cao et al. [38] found a parvocellular response and no koniocellular response whilst Zele et al. [329] found a koniocellular response and no parvocellular response. However, attention should be paid to the distinctions in experimental goals and procedures. Cao et al. [38] probe the long term adaptive effects of differing levels of melanopic activation upon white settings, whereas Zele et al. [330] probe the equivalent appearance of melanopsin in terms of cone-based percepts. It is possible that this distinction is the root of the inconsistency.

**Horiguchi et al. [148]** found that “[i]n the periphery, at high photopic levels,

---

11 Albeit inverted: S-OFF L+M-ON, rather than the traditional S-ON L+M-OFF [143].

12 Though the matching values for this pathway aren’t actually reported. One assumes they did not exhibit a trend and/or fell below some meaningful threshold.
human sensitivity is not accurately explained by absorptions in only three types of cone photopigments [and requires a fourth]”. This work focuses on discrimination thresholds rather than attempting to understand a direct perceptual correlate to melanopic stimulation. They conclude that “[t]he most likely hypothesis is that in healthy human subjects melanopsin absorptions influence visibility.”

Very recent results from Vincent et al. [309, 310] report that, contrary to expectations (based on the work of Allen et al. [2]), "sensitivity to flicker directed at the cones was not altered by adaptation to a steady field of substantially higher melanopic content". This appears to rule out the hypothesis that melanopsin activation alters gain (aka adaptation) at an early retinal level for cone-based signals.

Though all the above results are exciting, there are a number of methodological and theoretical areas which require development. The standard methodology used in these experiments is that of ‘silent substitution’ [95, 166, 280] where careful tuning of the spectrum is used (see the topic of 'metameric blacks’ [59, 304, 305, 307]) to generate signals which are only visible to the chosen receptor-type in theory, in this case melanopsin-expressing ipRGCs.

However, due to variability in observer spectral sensitivities, and limits on the level of stimulus control, it is likely that there will be a small amount of unintended stimulation of non-target cell groups. Spitschan et al. [281] refer to this as 'splatter’ (“the expected amount of contrast on nominally silenced photoreceptor classes for a given modulation around a given background”). A suggested control condition is to generate contrast for the nominally silenced cell groups at the level predicted from modelling.

A further concern is that horizontal cell feedback, or other intraretinal retrograde signalling, could result in nominally silenced cone populations still producing an output [166].

### 2.2.4 Value for colour constancy

In this section I shall outline the reasoning which suggests to me that there might be value in a melanopic input for attaining colour constancy.

Existing colour constancy algorithms fundamentally rely on the ability of cone-
2.2. Intrinsically Photosensitive Retinal Ganglion Cells (ipRGCs)

receptor-based signals to calibrate cone-receptor-based signals. In this context, by *calibrate* I mean *modify a raw signal in order to exclude unwanted elements of the signal, in order to improve the accuracy (and possibly also precision) of target signal measurement* where the *unwanted signal* would be variation in illumination, and the *target signal* would be either the SRF of a surface or some other identifier of the surface.

This general framework suffers from what I refer to as the issue of circularity in self-calibration\(^\text{13}\). Generally calibration is performed by characterising a sensor through measurement of an object where the ground truth is known (and/or a measurement from a trusted secondary sensor has been made of this object), then adjusting the properties of the sensor (either at the measurement stage or by implementing a post-processing stage) such that a measurement of the known object results in the expected values. In the case outlined above (cone-receptor-based signals calibrating cone-receptor-based signals), there is no ground truth object, and no secondary sensor, and thus calibration in the traditional sense fundamentally cannot be performed.

If only relative signals are of importance (as opposed to absolute value measurements), then an uncalibrated system might achieve satisfactory stability through the use of measurements taken over time or over space from a single sensor.

A melanopic signal could represent a secondary signal, and the properties of ipRGCs seem to make them well-suited for making measurements where the ambient illumination is preferentially detected over transient SRFs. Particular properties will be discussed below.

If one’s goal was to design a sensor which measured the ambient illumination upon a scene (and was only minimally perturbed by surface reflectances) it might be wise to limit both the spatial and temporal resolution relative to sensors which may be best for measuring surfaces, since the lighting on a scene generally operates at spatial and temporal frequencies which are both lower than surface variation within a scene, particularly when the scene is not viewed from a static position but

\(^{13}\)I suspect that this issue has been discussed in other fields but I have been unable to find such discussion as of yet.
from a constantly changing vantage (such as is the case with human vision, where
the observer is moving their body, face direction, and gaze direction regularly). It
would also be ideal if the secondary sensor was not strongly adaptive, as this would
allow for a more concrete relationship between stimuli and response. IpRGCs
exhibit all of the above properties.

There are multiple sites at which a melanopsin-based calibration could occur.
The synaptic connections to ipRGCs from rods and cones could allow a melanopsin-
dependent transform to be performed at the RGC stage before signals are output.
Alternatively, the intraretinal outputs from ipRGCs could allow for modification
of signals before reception by traditional RGCs. Finally, calibration could occur at
any higher post-retinal location assuming that melanopic and cone-based signals
could be reconstructed at that point.
2.3 Museum Lighting

“Museums and art galleries collect, preserve, and display natural artifacts and/or examples of human achievement and analyze their impact on the world and the universe around us. Effective exhibit lighting must balance exhibition and conservation needs and enrich the museum experience.”

IES RP-30-96 Museum and Art Gallery Lighting: A Recommended Practice [157]

Lighting in museums is required to satisfy multiple criteria; perhaps the least contestable requirement being that the lighting illuminate objects such that they are suitably visible to museum visitors. Also of utmost importance in most museum settings is that the lighting does not have an unreasonably damaging effect upon the objects or environment, be this through direct photodegradation or as a result of heat transfer. Further to these requirements, an increased or optimal visual quality is generally desirable, although what this represents or how to achieve it is generally ambiguous.

In sweeping terms, all electromagnetic radiation (visible and non-visible) damages objects, and more radiation damages objects proportionally more so. Thus the question becomes: how little light can we use to illuminate objects such that they’re visible to the extent required? The inverse form, sometimes used on the assumption that more light always represents an increase in observer satisfaction/pleasure is: ‘how much light can we use so that only x damage occurs over y time’.

Industry guidance documents provide advice on how to manage lighting to best address the above requirements and many other additional specific requirements through the recommendation of procedure and provision of target figures for quantitative variables. Ultraviolet (UV) radiation, being of no visual benefit but having potential to harm, is now excluded from gallery spaces as an industry standard.
2.3.1 Damage Functions

The key reference on this subject is CIE 157:2004 [49], and valuable talks on the subject were given at the recent Museum Lighting Symposium & Workshops [253] (which the author helped to organise), and have been made freely available online\textsuperscript{14}.

In heritage science ‘damage functions’ are “functions of unacceptable change, dependent on agents of change” [290]. The goal of damage functions in heritage lighting engineering is to give a quantitative means by which to predict the amount of damage caused to a prototypical object by a given light source, and to assist in limiting such damage. They generally follow the logic that radiation of lower wavelength is likely to cause more damage to objects.

Harrison [136] is generally acknowledged as the first to suggest such a function, but he himself acknowledges that it had “long been established that the shorter the wavelength (visible yellow, green, blue, violet and invisible UV being progressively shorter) the more photochemically potent will be such radiant energy, provided such energy is actually absorbed”. Harrison [136, p.9] defined the ‘radiation hazard associated with a light source’ as:

\[
\sum_{0}^{\infty} H_\lambda D_\lambda \Delta \lambda / \sum_{0}^{\infty} H_\lambda \gamma_\lambda \Delta \lambda
\]

where \(H_\lambda\) is the spectral irradiance, \(D_\lambda\) is the ‘Relative Damage Factor’ (which is extrapolated from the data collected shortly prior to Harrison’s own report by the National Bureau of Standards [235], and shown in Figure 2.22), and \(\gamma_\lambda\) is the CIE 1924 photopic \(V_\lambda\) luminosity function. The resulting value would describe the amount of damage expected from a light source, normalised by its luminance.

There has been extended scepticism about the utility of damage functions in general, with the argument being that no one damage function could represent the vast range and complexities of real materials. Thomson [297, p. 178] wrote that “for more fugitive materials … the figure for visible radiation would be higher. On the other hand … the fastest dyes are probably affected only by UV. Thus it can be

\textsuperscript{14}See in particular the talks by David Saunders (https://www.youtube.com/watch?v=H4d0qH0Bcl&t) and Stefan Michalski (https://www.youtube.com/watch?v=XUY9biLQqw).
Figure 2.22: Harrison’s [136] damage function (\(D_\lambda\)), and luminous efficacy (\(\gamma_\lambda\)), alongside the Declaration of Independence data [235] from which it was extrapolated (re-normalised to match scale).

seen that no single figure can be given for damage versus wavelength”.

Criticism was aimed at Harrison’s specific damage function due to its derivation from such a small and minimally representative dataset - data was ‘extended’ from the 5 datapoints measured by the National Bureau of Standards [235] in their investigations of how to best care for the Declaration of Independence\(^{15}\), and was derived from the study of ‘low-grade paper’, which cannot to said to represent the average museum item\(^{16}\).

CIE 157:2004 [49] notes that whilst Harrison’s proposal failed to gain acceptance as the procedure for comparing the damage potential of different types of light sources, it did convince people of the ills of UV, with the result that daylight

\(^{15}\)It is a curiosity that these minimal figures would not in fact have been much use to those planning the care for the declaration, since in the report it is noted that “The deterioration of animal parchment is not as rapid as that of the low-grade paper for which the damage factors were determined” [235], and the Declaration of Independence is written on animal parchment.

\(^{16}\)Though of course no material truly can!
was subsequently eliminated from many galleries.

Following Cuttle [64], who noted that Harrison’s damage function could be well fit by an inverted logarithmic function, with parameters controlling the slope and normalisation point of the function, CIE 157:2004 provided the following equation:

\[ s(\lambda)_{\text{dm,rel}} = \exp[-b(\lambda - n)] \]  \tag{2.13}

where differing values of \( b \) for 5 categories of item are provided (Table 2.1), \( n \) is the normalisation value (CIE 157:2004 uses a value of 300), and the \( s(\lambda)_{\text{dm,rel}} \) function is the estimated action spectrum for each category. \( s(\lambda)_{\text{dm,rel}} \) would be substituted into Equation 2.12 for \( D_\lambda \). CIE 157:2004 also provides values of \( H_{s,\text{dm}} \) which indicate the susceptibility of each group of materials to damage (where damage is considered as colour change).

<table>
<thead>
<tr>
<th>Group</th>
<th>Samples</th>
<th>( H_{s,\text{dm}} ) (W h/m²)</th>
<th>( b )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Low-grade paper</td>
<td>5</td>
<td>0.038</td>
</tr>
<tr>
<td>b</td>
<td>Rag paper</td>
<td>1200</td>
<td>0.0125</td>
</tr>
<tr>
<td>c</td>
<td>Oil paints on canvas</td>
<td>850</td>
<td>0.0115</td>
</tr>
<tr>
<td>d</td>
<td>Textiles</td>
<td>290</td>
<td>0.0100</td>
</tr>
<tr>
<td>e</td>
<td>Water colours on rag paper</td>
<td>175</td>
<td>0.0115</td>
</tr>
</tbody>
</table>

Table 2.1: Table reproduced from CIE 157:2004 [49], showing the values for \( H_{s,\text{dm}} \) (threshold effective radiant exposure) and \( b \) for various categories. Note: the source for this data is not particularly clear; it is listed as ‘The Berlin researchers’, which is assumed to follow the references: CIE [47] (which I have been unable to access), Hilbert et al. [146] which provides the values of \( H_{s,\text{dm}} \), and Krochmann [179] which appears to provide equations similar to Equation 2.13 but no specific values for \( b \).

The more general criticism that damage functions will never be able to represent all museum objects is a valid concern, and can be well illustrated with the following logic: museums own objects of many different colours, different colours arise from different reflectance properties, different reflectance properties mean different wavelengths are absorbed, and damage can only occur when radiation is absorbed. Thus it follows that one would expect two objects of different colours to have different damage functions.
The classic study on how reflectance properties relate to damage is that of Saunders and Kirby [264]. They exposed a number of pigments to a range of wavelengths and measured the resulting damage. Cuttle and Ne’eman [67] later replotted the data from this study (see Figure 2.23), highlighting the apparent joint contributions of spectral reflectance and a general damage function to the individual damage functions. CIE 157:2004 notes however that this correspondence is not perfect or easily modellable, and that “a workable system for characterising action spectra for colorants, including pigments and dyes, remains an unattained goal”. Recent studied have added new data (see esp. Villmann and Weickhardt [308]), and it is hoped that a general understanding may at some point be reached.

However, it is the opinion of this author that this argument is an exercise in artificial futility; whilst we may not be able to model the individual damage functions for every object in a museum, we may at least use one which has some bearing on the damage function, rather than the one which is implicitly used by museums currently - the CIE 1924 luminosity function (\(\gamma\) of Figure 2.22), which relates to the sensitivity of the human eye rather than any type of object. Cuttle [64] puts it well: “The argument, then, is not whether we have a [damage] function which is correct, but whether we can improve usefully upon the likely reliability of the present system”. It is depressing that this was said in 1988 and yet little seems to have changed in practice (See Chapter 3).

In the case where specific objects/pigments of interest can be identified, and their individual damage functions calculated, there is valuable research to draw on regarding methods for optimising light sources to minimise damage, initiated by Miller [217] and developed by many others [65, 74, 76, 77, 85–88, 302]. The fading of lead chromate in the paintings of Vincent van Gogh has captured public attention [191] and has resulted in multiple studies looking at ways to optimise the illuminant for this one particular material [197, 222].

With modern computation, and access to datasets, it becomes relatively easy to calculate a value of damage index (DI). Figure 2.24 shows the results for such computations for 401 illuminants and light sources (as per Equation 2.12, using
Figure 2.23: The original data for this figure is from Saunders and Kirby [264], later replotted by Cuttle and Ne’eman [67], and reproduced (with the addition of the Berlin functions) by the authors of CIE 157:2004 [49]. Caption from CIE 157:2004: “Spectral absorptance (solid line) and relative spectral responsivity (broken line) for artist’s pigments … The dotted lines show relative spectral sensitivities normalised at 400nm based on the Berlin relative spectral responsivity function.” It seems that in this context the CIE document uses the term ‘spectral responsivity’ to refer to measured change, and ‘spectral sensitivity’ to refer to predictions of damage, though this distinction is not particularly clear.

A damage factor computed as per Equation 2.13 with a value of 0.0115 for $b$, but further normalised such that Illuminant A has a reference value of 1)\textsuperscript{17}. It can be seen that whilst most illuminants cluster around 1 there is a broad range. It should be remembered that these illuminants would in no way indicate their relative damage index to an observer or a purchaser, unless one went to the effort to look up or measure the SPD and compute the damage factor. A careful or careless choice in this respect could easily double or halve the amount of time an item could be exhibited before succumbing to terminal damage.

\textsuperscript{17}The code to reproduce this is available from: https://github.com/da5nsy/DamageIndex
Figure 2.24: The values of damage index (DI) for the 401 illuminants used by Houser et al. [150], available via Psychophysics Toolbox Version 3 (PTB) as ‘spd_houser’, normalised such that CIE Illuminant A receives a reference value of 1.
2.3.2 Visitor Requirements of Museum Lighting

The visual requirements of museum visitors are likely to depend upon a large number of variables, both intrinsic and extrinsic to the visitor (e.g. intrinsic - age, cultural background, goals for the museum visit and extrinsic: luminance, lighting distribution, CCT, CRI and flicker properties of lighting). Some of these factors have been independently studied in a museum context, and for others it is likely that findings in other environments could generalise such that they could be used to inform decisions regarding museum lighting.

Traditionally, the principal manner in which museum professionals sought to limit damage was through setting a maximum luminance level. The implicit assumptions in this process are twofold; firstly: that damage will increase with increased luminance. This was a fairer assumption when tungsten was the only type of lighting technology, but as other lighting technologies with different SPDs have been introduced this assumption has become less accurate (see Section 2.3.1: Damage Functions). The second implicit assumption is that viewers will prefer higher luminance environments.

The classic study on this second assumption, performed in a mock-museum environment is that of Loe et al. [194]. This research regards the display of oil and watercolour paintings specifically. In this study Loe et al. examined three variables: painting illuminance, light source (different technologies) and light distribution. Following the construction of a mock up gallery space, observers were asked to view paintings of various types under a range of illuminations, varying in ‘painting illuminance, light source and light distribution within the gallery space’ and report upon semantic scales their perceptions. The results which informed the 200 lux recommendation stem from only the first variable, painting illuminance. Here it was found (as shown in Figure 2.25) that for factors christened ‘discrimination’ and ‘quality evaluation’ (distilled from factor analysis of the original semantic data) there was ‘a steep rise in discrimination and quality assessment until and illuminance of approximately 200 lux is reached: above this illuminance the rate of increase in reduced.’ This conclusion has had substantial impact in setting guidelines and future
thinking was that a minimum of 200 lux was required to 'give visual satisfaction', however it can be seen from Figure 2.25 that the data is sparse, noisy, dependent upon brand of light source and doesn’t show a particularly strong effect of 200 lux in particular. Further, only a small number of different luminances were sampled, and it is quite possible that the results are at the mercy of several types of bias [116]. The value of 200 lux was subsequently used in Thomson [297], which has informed a great deal of subsequent thinking on the topic.

Figure 2.25: Illuminance vs a factor which is thought to indicate 'quality', for three different brands of illuminant, reproduced from Loe et al. [194].

There has been extensive research on the visitor requirements in terms of CCT and CRI and these shall be covered in separate sections.

In terms of a holistic approach to museum lighting research (considering more than a single or small number of isolated variables), there has been a great deal of work which deals with the visitor experience in a broad and cultured fashion from
the lofty vantage of museum studies (such as Falk et al. [99] and Shapiro and Kemp [268]), but rarely do the physical practicalities such as lighting get a mention.

The notable exception, where lighting as it relates to the visitor experience is considered, is the work of Kesner in the USA in the early nineteen-nineties [169–173].

Kesner’s studies comprised self-reporting surveys, and concluded that colour accuracy was the highest requirement and that ‘richness’ of colour was the lowest priority:

“Artifact appearance, particularly clarity of artefact form and accuracy of artifact color, is the most important visitor need. Although visual impression, specifically acceptable gallery brightness and rich artefact color, is least important among the factors, it too rated highly important.”

Kesner [169]

2.3.3 Museum Lighting Specification Guidance Documents

For a historical perspective on museum lighting guidance see Druzik and Eshøj [82].

Five museum lighting guidance documents, thought to represent the most commonly referred to documents in the field (informed by the interviews reported in Chapter 3), were reviewed.

The goals of this section are:

1. To explore how museum professionals specify lighting, by understanding the tools and guidance which are available.

2. To enquire as to what the guidance actually is, in terms of what subjects are covered and what the guidelines actually are.

3. To question what these guidelines are based upon. How are the results of scientific study utilised in the production of these documents?

\[18\] A cautious interpretation of these results is advised, considering that there were high average scores for each category.
2.3. Museum Lighting

Limiting the damage to museum objects by photodegradation is most often the responsibility of the preventive conservator within a museum, or the person holding a role which encompasses this role in the case of smaller institutions. It is therefore essential that the people in these roles have access to standardised and validated advice on how to approach the subject. Several overviews of the subject have been written and I shall aim to introduce the most prominent here, focusing on their aims, scope and differences/similarities.

I shall operate a biased interest towards the recommendations regarding colour rendering, CCT, and illuminance level. Whilst the first two are the immediate area of study within this project, the third (illuminance) is of interest for multiple reasons. Firstly, it is the most prominent area in which lighting guidance is provided, on the assumption that there exists a correlation between illuminance and damage, and secondly because of the unit of specification - lux, which is generated using a function of wavelength designed to provide a correlate of brightness to humans. As a human-based function, it is within the scope of interest to this project.

There shall be an active occlusion of any advice, no matter how interesting, pertaining to subjects other than lighting and to areas of lighting guidance which do not fall into the above remit, such as directionality, UV/IR damage, advice relating to specific technologies and areas of discussion such as cost calculations or warranty considerations.

Firstly, a general note regards the mind-set of those providing recommendations: conservation recommendations provided to museums, which might be presumed to be concerned with a method for limiting damage, are generally in no way informed by the sensitivity of a prototypical object, nor how light might act upon it, but rather it is concerned with maintaining a minimal acceptable light level for an observer to view objects under.

This is based on the argument as follows: all light is damaging, but required in order for visitors to see objects. Damage by light acts in a roughly reciprocal manner, such that a small amount of light over a long time period might be made to do equal damage as a large amount of light over a short period. Considering the
multiple aims of museums; (1) to display objects, but also (2) preserve them such that they may be displayed to future generations, it seems desirable to find the minimal amount of light that satisfies the first aim, such that the ability to deliver on the second aim is maximised. Thus many of the recommendations discussed here are actually concerned with the ability of a viewer to extract visual information. This point is not always entirely clear, and it is my suspicion that conservators are sometimes lulled into believing that there is something special about the specific values recommended regards their ability to ‘avoid’ damage.

One clear exception to this, which will be discussed in further detail, is CIE 157:2004 [49] which considers damage functions of specific materials. These deal specifically with the degradation of objects, with less regard to how objects might appear. Other works which deal with damage functions either generally or for specific types of objects fall into this exception also.

The common language most regularly employed in both of these approaches is the term of ‘lux’, though this is a contentious issue with some arguing that it correlates well enough with damage potential, and others exploring the use of damage functions. ‘Lux’ relates loosely to the human perception of brightness, but does not consider any type of material absorbance, reflectance or damage function. The historical background to this precedent appears to be as follows; whilst technological ability to vary the spectral power distribution of a light source was limited (whilst tungsten was the dominant source of illumination) it could be assumed that there was a predictable relationship between lux and radiometric spectral power distribution, such that any two lights with the same lux level would cause the same level of damage to any specific object, and thus providing damage minimisation advice could be done using the language of lux, which was pre-established considering the original approach outlined above. This approach is now less appropriate, considering the increased variability in spectral power distribution provided by the introduction of LED technology, and it is reasonable to assume that in the future additional lighting technologies (or variations of existing technologies) might be introduced, with again fundamentally different types of spectral power
distribution.

This field may be divided into publications which relay original research, generally in the form of journal articles and conference proceedings, and longer publications which aim to provide an authoritative voice on the subject (often including references to the aforementioned journals). I shall cover principally these longer documents, since my priority interest in this section is the advice which is currently provided and the methods in which this advice is conveyed. The question of how much conservators rely on guidance documents vs referring to ongoing research is an interesting question in of itself, and shall be covered in my discussion of the interviews conducted with current museum professionals (See Chapter 3).

The guidelines chosen for review were:

- The Museum Environment [298]
- CIE 157:2004 Control of Damage to Museum Objects by Optical Radiation [49]
- Guidelines for Selecting Solid-State Lighting for Museums [83]
- SLL LG8: Lighting for museums and art galleries [46]
- IES RP-30-96 Museum and Art Gallery Lighting: A Recommended Practice [157]^{19}.

In summary:

- Recommended values were provided for:
  - \textit{Lux}: various, dependent on sensitivity, generally based on figures from the study of visual preference by Loe et al. [194] via Thomson [297].
  - \textit{Ra}: various, most frequently >80, generally with no experimental basis referenced or justification for this exact figure.
  - \textit{CCT}: various, based implicitly on Kruithof [180] or on such ideas found empirically.

^{19}This has since been usurped by ANSI/IES RP-30-17 [159].
Most also suggested visual inspection as a valid means of assessment.

There was often blurring between recommendations concerned with visual appearance and those concerned with conservation.

2.3.3.1 ‘The Museum Environment’

‘The Museum Environment’, first published in 1978 [297], with a popular second edition published in 1986 [298] (I shall hereon be referring to this later edition), appears to be one of the most frequently consulted resources on the subject of lighting for practising museum professionals. Whilst the book encompasses a great many subjects aside from lighting, two chapters are set aside for the subject of lighting specifically, covering a large range of topics within the scope of museum lighting. In Druzik’s overview of museum lighting specification [82] he notes that ‘up until the first edition of The Museum Environment in 1978, no one had written a book on preventive conservation in museums that was comprehensive, yet clear enough for scientists and conservators to use with nearly equal ease.’

The section of Thomson’s book most often referred to in my experience has been his guidance on recommended exposure for different object types. A table from this section is reproduced here as Figure 2.26, and details the types of objects which might most readily be assumed to fall into groups of like sensitivities. The recommended maximum illuminance values stem most heavily from experiments performed by Loe et al. [194] in the years immediately previous to the publication of this second edition. It is noted that these recommendations are a revision upwards from the first edition, where 50/150 lux was recommended in place of 50/200 lux.

Loe et al.’s work concludes with the recommendation that “preferred artificial lighting conditions for viewing works of art are a painting illuminance approaching 200 lux provided from lamps with good colour rendering characteristics, e.g. a CIE $R_a$ index of >85 and a large gamut area”. It is of particular interest to me that Thomson neglected to quote the final part of this recommendation regarding the specific CIE $R_a$ and note regards gamut area.

On the topic of colour rendering, Thomson goes into great depths to explain
Figure 2.26: Table 2 of Thomson [298, p. 23], describing recommended maximum illuminances for 2 groups of objects.

The topic and the means available for calculating various indices, spending a disproportionate amount of time (compared to what I understand was normal for the time) on the topic of Crawford’s ‘band method’ [62, 63] for colour rendering index calculation amongst other things.

For the most part he neglects to make recommendations for specific target values, only making the rough and unsupported assertion in passing:

"Recommendation for good colour rendering.

\[ CIE \ R_a \ about \ 90 \ or \ better, \]

\[ R_w \ (\text{worst } R) \ about \ 80 \ or \ better, \]

and Crawford Class A, B or C."

No research is referenced to support this recommendation. The inclusion of \( R_w \) (the lowest value of \( R_i \), the rendering index for an individual TCS before averaging and scaling is applied to produce \( R_a \)) is notable due to its lack of consideration in other documents.
Still on the topic of colour rendering, of additional note is Thomson’s general introduction to the subject, where he frames the conversation in a manner which presents colour rendering as a measure of the colour changes to which the human visual system is unable to adapt to, opposite changes induced by changes in CCT to which the human visual system is generally able to adapt easily to. Thomson also provides an accessible, and often quoted summary of the underpinnings of how the CIE General Colour Rendering Index operates in principle:

“Adapt our eyes to the illuminant under test.
Look at a set of representative objects under it and accurately memorise their colours.
Adapt to the reference illuminant.
Look at the same objects under this second illuminant and compare the colours to the colours in our memory.”

Thomson’s comments on the selection of CCT are minimal, not referring to any specific CCT in his summary of specifications [p. 268]. The only note dealing directly with the specification of CCT can be found on page 25, and refers only specifically to the conversation of what CCT to select for those environments which for conservation purposes need to be lit at particularly low illuminance levels.

“The ‘coolness’…of daylight when it has been reduced to 50 lux often gives the impression of gloom, especially when it is highly diffused. No one knows how deeply it has been built into our systems, but ever since our first ancestors sat around fires, and later used oil lamps and candles, the human race has been accustomed to ‘warm’ light in the home after dark. As a result the warm 50 lux from tungsten lamps appears to be brighter, and certainly more cheerful, than 50 lux of diffused daylight.
For the same reason warm rather than cool fluorescent lamps should be chosen for 50 lux situations.”

It is assumed that the views expressed above are grown from empirical observation. They mirror the standpoint of other publications which refer to the work of Kruithof [180] and the derived ‘Kruithof curve’ (see Section 2.3.4).
The bulk of pages 49-51 concern CCT, if one includes the associated discussion of chromatic adaptation and colour constancy, but there is no clear link as to how the author suggests that this theory should be considered in practical application, nor any concrete recommendations for target figures.

A final note in reference to this text goes to the discussion of whether or not to consider the type of lighting for which the artist intended an artwork to be displayed under, or that under which it was originally created. At the risk of quoting half the book I include one final passage which I think particularly pertinent to the area of research to be undertaken in this project, with which I shall conclude my discussion of this publication:

“I think one would be correct to suppose that artists have always assumed that their creations would be viewed in a variety of situations, not all lit ideally, and have designed their work accordingly. Even the Impressionists and others who made a point of completing their canvasses in the open air did not expect them to be so viewed. When De La Tour painted a candle-lit scene he painted it in such a way that the scene would look candle-lit under any reasonable lighting (for a contrary view see Weale [[317]]).

But it could also be said that, however robust the work of art, it will look better in some lighting situations than in others, and so we should bend our efforts to finding the best possible situation. Within the limitations of the museum one cannot but agree, provided there is indeed a consensus of perceptive opinion on what is best, and provided that the damage caused by light is kept under control.

There is certainly no mathematical treatment whereby we can equate the viewer’s gain against the exhibit’s loss. However there has been considerable research on the visual process as it is affected by the lighting, and Brommelle [[33]] has carefully related the experimental work to the museum problem.”
2.3.3.2 CIE 157:2004 Control of Damage to Museum Objects by Optical Radiation

This document is a CIE technical report, prepared by CIE Technical Committee 3-22 of Division 3 "Interior Environment and Lighting Design".

To set the context for the discussion of this CIE technical report, first a note to draw attention to the time period during which it was drawn up, since to the future reader there might be some ambiguity as to what the state of the art was at this point. Whilst it is now at the time of writing in 2016 common to find museums using LEDs to display their objects, this was not the case in 2004. It is noted within the text that LEDs are ‘are not of suitably high colour quality for museum use at present, but have future potential as very low UV power sources.’ The increased use of LEDs does not invalidate the contents of this document, but it is worth considering that it was prepared with pretext to address any LED specific issues.

As previously mentioned, this document takes a different approach to the problem of limiting light induced damage in museums, focusing rather on the process of considering the optimal spectral power distributions as opposed to limiting the light levels wholesale. For example, the closest the document comes to making light level recommendations in the manner of ‘The Museum Environment’ [298] is in the tabulation of Mlx hour values for predicted noticeable fade. This approach leaves the decision of actual lighting levels to the museum professionals, with the question being 'How long do I want this object to last before a noticeable fade has occurred?'

The document is in some ways an endorsement and extension of the approach taken by Harrison [136] (as discussed in Section 2.3.1), in which the concept of a damage function was introduced, this being "an action spectrum that defines the relative spectral responsivity of a receiving material". The authors of this document note that the work did not originally gain traction due to conservators’ scepticism that a single function could represent the vast range of potential museum object (still an intractable problem). Interest was further lessened by the fact that the original work refers only to one very specific type of material - low grade paper.
However, the authors of this document argue that the fundamental idea is sound; a damage function (so long as it is genuinely representative to an extent) could be a valuable tool in assessing the appropriateness of different light sources.

The authors then pull reference from further studies focused on a range of other materials, and find low grade paper is actually an outlier in terms of wavelength sensitivity, with many other materials able to be grouped together in type of dependency (if not like sensitivity). Following Cuttle’s work [64] to describe the pre-existing damage functions as simple mathematical relationships, comparison between the variety of newly created damage functions was possible.

Regarding CCT, the report states:

“While the variations of spectral responsivity for individual materials, particularly pigments, remains problematic, the overall tendency for responsivity to increase at shorter wavelengths is reasonably well defined. It has been shown that there is a general effect for the relative damage potential to increase as colour temperature increases [Cuttle [64]]

This falls short of offering recommendations for CCT choice, remaining impartially scientific. However, an included table, reproduced below, makes it clear that the lower the CCT, the lower the potential for damage (for a set SPD ‘type’).

This section concludes with the introduction of two related topics, again with significance for this project. The first regards the work of Miller [217] in which the concept of Reflected Energy Matching (REM) is introduced, the second is of the work by Thornton [299] in which he describes the technique of using ‘prime colour’ light sources.

REM in the current context refers to a process of illuminating an object preferentially with illumination of wavelengths which it is known to reflect. The premise of this is: absorbed radiation is visually unimportant, yet conversely the radiation most likely to damage an object. Thus an ideal situation would be to illuminate an object with radiation which it entirely reflects. The immediate limitations of this process are that this process would be object specific and potentially problematic for correct colour rendering. This procedure was suggested to be implemented
using fibre optics and filter systems, it would likely be much more approachable with the advent of LED technologies which allow certain spectral flexibility.

A note in the text regards the potential for REM to increase the saturation of colours relates that this ‘raises questions concerning the ethics of modifying the apparent colours of displayed objects.’

In the final section of this document (written with the aim to ‘give recommendations for lighting in museums’) very little advice is given in the form of specific figures to aim for, rather a decision appears to have been made that it was preferable to detail the areas that the authors considered worthy of attention, and to lay out the arguments for deliberation. For example, on the topic of CCT choice reference is made to the previous discussion which details the deleterious potential of illumination at different CCTs, but it is clearly stated that ‘[w]here the viewing conditions call for moderate or high colour temperature lighting, conservation concerns should not override design objectives for the display.’ This seems to be supported by the later statement that ‘[i]t is thoroughly bad policy to place an object on display, where it inevitably will suffer some damage, and to fail to present it adequately.’ It is further noted that low CCTs may ‘judged unsatisfactory’ when it is used in combination with natural illumination.

<table>
<thead>
<tr>
<th>Colour temperature of a Planckian source</th>
<th>Relative damage potential</th>
<th>D series source</th>
<th>Relative damage potential</th>
</tr>
</thead>
<tbody>
<tr>
<td>2500 K</td>
<td>0.92</td>
<td>D55</td>
<td>1.63</td>
</tr>
<tr>
<td>3000 K</td>
<td>1.04</td>
<td>D65</td>
<td>1.87</td>
</tr>
<tr>
<td>3500 K</td>
<td>1.20</td>
<td>D75</td>
<td>2.07</td>
</tr>
<tr>
<td>4000 K</td>
<td>1.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4500 K</td>
<td>1.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5000 K</td>
<td>1.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5500 K</td>
<td>1.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6000 K</td>
<td>2.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6500 K</td>
<td>2.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7000 K</td>
<td>2.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7500 K</td>
<td>2.40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.27: Table 2.2 of CIE 157:2004 [49].
With marked separation, the previously discussed work of Loe et al. [194] is referred to, in the context of providing guidance on a lux value that is ‘generally sufficient to provide for adequate visibility’.

There is a noteworthy section on the meaningful distinction between illuminance and irradiance: ‘It needs to be recognised that illuminance is not a reliable alternative measure [for irradiance], as it represents the density of luminous flux, being radiant flux evaluated according to a typical human visual response, defined by the photopic spectral luminous efficiency function V(λ). Not only does illuminance take no account of irradiance outside the visible spectrum, but also radiant flux within the visible spectrum is weighted according to its relative visual effect, which is not related to its damage effect.’

2.3.3.3 CGI/Ge/t_ty Guidelines for Selecting Solid-State Lighting for Museums

This guidance document [83], produced by the Canadian Conservation Institute and The Getty Conservation Institute, provides guidance for museum professionals to select LED lighting. As such, the guide includes an introduction to the technological theory of LEDs, advice on how to consider the requirements of museum lighting, and a practical guide to selecting lighting. It is the only document here considered which discusses solely LED technology.

Compared to the other documents considered here, there is also a focus on the longevity of systems, particularly in respect to potential for colour change, and advice on the type of information that should form a warranty agreement between supplier and end user.

A recurring piece of advice throughout the document is that a lighting specifier should always view lighting in person before making a significant order. The implicit, and sometimes explicit subtext here is twofold; firstly, that the specifications available for lighting are insufficient to describe the visual appearance of lighting, and secondly, that the quality of museum lighting (the ability of lighting to fulfil predetermined requirements) is visually assessable. An extension of this second point would be to consider the visual requirements in museums as being solely or
principally defined by a general preference for appearance under lighting.

Following this notion, CIE $R_a$ is referred to as ‘imperfect’ and ‘misunderstood’, and it is suggested that it could be used as a ‘secondary consideration’. There are several different specific values of CIE $R_a$ recommended at different points within the document.

“It is generally agreed that a CRI above 85 is suitable”

“To illuminate areas with a more utilitarian [function] such as machinery, many science exhibits, food services, hallways, educational activities, etc. settle on a color rendering index (CRI) above 80. When color matching may be more an attentive activity such as viewing art, ethnography, some natural history collections exhibits, etc. select LEDs with a CRI above 90. However, because CRI is an imperfect metric, CRI should be considered a target, not a firm criterion.”

“There is no international museum standard on what is or is not an “acceptable” CRI, but the Canadian Conservation Institute (CCI) recommends a minimum of 85. Many museums specify greater than 90.”

As is perhaps apparent in the quotes above, this document has the feel of friendly advice rather than an official guidance document. As such, it has particular value for this project as a perhaps more revealing account of the advice which actually circulates in the field. For example, one piece of advice which I haven’t seen anywhere else in official literature, but which as I understand it is relatively common in general parlance is that provided during the overview on page 23: ‘Check color rendering on your own skin’. This advice is incongruous with the museum lighting advice which prioritises objectivity and impartiality, as this implicitly suggests looking for a ‘pleasing’ appearance.

The document makes subtle references to the Colour Quality Scale (CQS) [11, 71, 72, 240] and Gamut Area Index (GAI) [257]; saying that in response to limitations with CIE $R_a$ ‘at least two other color metrics have been proposed recently’, followed by references which relate to the aforementioned respectively. The
document also makes two references specifically to CIE $R_9$ values, firstly as part of the information on the ENERGY STAR program (where it is quoted alongside other lamp specifications) and once in the Appendix (where visual demonstrations of induced colour shifts are provided for the CQS and CIE test colour samples. Interestingly, a scale for CIE $R_9$ is described, where: ‘CIE $R_9 = 0-49$ means it renders red hues well. When $R_9$ is $50-74$ it is very good. $R_9$ above $75$ is considered excellent.’ As with many items in this document, no reference as to the source of this information is provided.

Regards the question of CCT within museum lighting, the guidance offers a Kruithof-ian approach, though without naming it as such, recommending warmer light for lower light levels and colder light for higher light levels.

“With low light levels, as in museums, viewers tend to prefer warmer light similar to that of incandescent lamps, e.g. the 2800K of standard incandescent lamps, or the slightly higher 3000K of quartz halogen incandescent lamps. As illumination increases to several thousand lux, preference is for cooler light, 5000K or higher.”

It is noted than an exception might be made in the situation where artificial lighting is employed to augment natural illumination. This said, elsewhere in the same document the reader is advised to ‘avoid higher color temperatures for light sensitive materials as these LEDs may have an unacceptably large peak in the “blue region” of the spectrum.’

Also poorly referenced is the discussion of the 50 lux recommendation. Research is mentioned but not directly quoted. It is assumed that the research described as ‘in the 1980s’ is the previously discussed research of [194].

The final area of note within this report is that which considers the methods applicable to assessing museum lighting in situ. A survey completed at the Field Museum [231] is quoted in detail, the survey having been completed by museum staff and lighting practitioners as part of a GATEWAY program demonstration at the museum. This survey includes such questions as ‘The lighting product shows ________ of the subject colors accurately’ and was completed for a halogen
system and an LED system in the same space.

2.3.3.4 SLL LG8: Lighting for museums and art galleries

This document [46] is one of the more recent industry standards, and benefits from numerous practical examples.

CCT is mentioned only in relation to its use as a creative tool; ‘to have an effect on the mood of a space and the exhibits in it’, and to match with daylight in mixed illumination settings.

It is stated that CIE $R_a$ values above 90 are ‘considered to be very good’, and those below 80 ‘will normally not be appropriate for lighting exhibits in museums and galleries’.

A particularly poor description of colour rendering is given in an image caption, where the image shows three bowls of fruit running from de-saturated to highly saturated left-to-right; ‘Colour rendering index (CRI) is a general indicator of how ‘natural’ object colours will appear when illuminated by a particular light source. This shows an indication of the dull colours of an unacceptable CRI of 50 (left), moderate CRI of 70 (centre) and preferred CRI of 90 (right) with full vibrant colour range’ (bold mine).

The illuminance guidelines of Thomson [298] are used (50 lx and 200 lx), with an additional intermediary category added for ‘medium responsivity’ objects. No specific guidance is given on how to determine what category an item should be considered as.

A sensible caution is given regarding the consideration of unavoidable damage:

“*As any light exposure will cause damage, it is inevitable that a compromise will have to be struck between the needs of the viewer to see, interpret and enjoy the material and the latter’s long-term preservation. Using the least amount of light practicable remains the primary goal when working with light-sensitive materials.*”

“*The commonly recommended illumination levels of 50 and 200 lux are based on established practice and research into visual acuity and are not*
levels below which damage does not occur - a frequently misunderstood concept. Similarly, there is an erroneous belief that the filtering of ultraviolet (UV) radiation provides complete protection and makes the light ‘safe’. Removal of UV offers some protection for colours that are sensitive to UV, normally the more stable colours, and can help to limit structural changes. ”

Specifically regarding LED products, the guide suggests that side-by-side trials are essential in judging quality.

In summary, this is an accessible guide (in terms of level of expertise and writing style), which follows the same guidelines as history suggested. Information regarding a moderate number of different types of material may be useful to some readers.

2.3.3.5 IES RP-30-96 Museum and Art Gallery Lighting: A Recommended Practice

This guide [157] has been recently updated [159], thought I have not had opportunity to access the updated version.

The guide starts with the three rules of Michalski [216]:

“ The artifact should be visible when on display. There is no point causing a little damage (with insufficient light) for no purpose (the artifact cannot be seen).

The institution must decide how much light damage in how much time is acceptable, i.e., what lifetime is desirable.

The institution must acknowledge the sensitivity of each artifact, or group of artifacts, as accurately as possible. ”

Regarding colour of lighting (CCT and colour rendering) the guide states that ‘color should not change the look of an artifact [and that one should] assure that the artifacts illuminated appear as their makers intended’.

The guide has a valid description of colour rendering and suggests CRI values (presumed CIE $R_a$) of $> 80$. CCT is mentioned only as a creative choice.
Chapter 2. Literature Review

The guide notes the 30 lux is ‘required for colour perception’. This guide is thorough and well referenced.

2.3.4 CCT in Museums

CCT is often described as an important variable in museum lighting, but definite recommendations, in the rare cases that they are given, are generally based on nostalgia for the appearance of tungsten lighting, questionable research on human preference [117, 180] or very rough rules that predict that damage potential will be decreased if CCT is minimised [49]. Whilst some research appears to have found optimal CCTs for viewing artwork [102, 193, 233, 251, 266, 267, 303] results often have large inter-observer variability, context dependency and it is not uncommon for the headline findings of separate studies to be in contradiction.

Specifications often refer implicitly or explicitly to the findings of Kruithof [180], who found that at lower levels of illumination, lower CCTs were preferred, and that at higher levels of illumination higher CCTs were preferred (see Figure 2.28). Whilst this general trend seems to have anecdotal support, it is possible that there may be lighting-technology-based confounds, and recently researchers [306] (including a meta-study of multiple other examinations [117]) found there to be no substantive support for Kruithof’s findings.

The damage justification seems more substantive; following the application of damage factors as discussed in Section 2.3.1 the CIE published a report showing the varying the CCT of museum lighting could have a clear impact on the potential damage undergone by museum objects [49]. The key figure is reproduced here as Figure 2.29.

This line of reasoning relies heavily upon the applicability of damage functions to the specific materials in question. As discussed previously (Section 2.3.1) it seems that whilst the generalised damage functions account for some of the individual material damage functions, they do not fully do so. However, tentatively, they do seem to be a relatively good fit for the shared characteristics of different damage functions (at least far more so than $V(\lambda)$), and so it seems reasonable to use them where no other broad approximation for a large set of objects exists.
To verify the findings of the CIE, extend their findings to LEDs, and provide code for others to do similar research or even test their own lights/materials, a set of MATLAB functions\textsuperscript{20} have been written (by the author) which calculate DI values for arbitrary illuminants. This code has been used to produce Figure 2.30, which shows the CCTs and DIs of the 401 SPDs of Houser et al. \cite{150}. A $b$ value of 0.0115 corresponding to ‘oil paints on canvas’ and ‘water colours on rag paper’ (see Table 2.1) was used. It can be seen that there is a clear correlation between CCT and DI with higher CCTs being predicted to be relatively more damaging. If a damage function with a lower $b$ value (such as for ‘low-grade paper’ or ‘textiles’) had been used, the relationship between CCT and DI would be steeper.

\footnotesize\textsuperscript{20}https://github.com/da5nsy/DamageIndex/blob/c7851e27ca1b0915013d8723db04704b49b4085e/CalcDI.m
Figure 2.29: Figure reproduced from CIE 157:2004 [49, p.16], showing the relationship between CCT and relative damage potential for black body radiator illuminants, CIE source A and three CIE D-series illuminants.

The results of these computations show a clear relationship between CCT and DI. Considering that this is not currently taken advantage of in museum lighting, it seems as though there is a great potential for reducing damage whilst maintaining visitor visual satisfaction.

Figure 2.30 (on next page): The CCTs and DIs of the SPDs used by Houser et al. [150] (provided via personal communication, but now partially (without category information) available via PTB as ‘spd_houser’).
2.3.5 Colour Rendering Indices in Museums

The first chapter of Cuttle’s book ‘Light for Art’s Sake: Lighting for Artworks and Museum Displays’ [66] is titled ‘A philosophy for the presentation of art’. The choice of title here is apt but may surprise some readers, sounding more whimsical than might be expected of a serious subject attended by scientists and engineers. The reason it is so apt is because lighting is unavoidably a creative intervention\(^{21}\) which is to say that there is no lighting which is truly impartial, and no lighting which is truly ‘correct’ in the sense of being unequivocally superior to another. All lighting decisions require choices to be made, and whilst these choices can be wrapped up to appear as an optimisation problem where proximity to a particular solution is the goal, the problem always rests on the bedrock of a philosophy based decision.

In the above mentioned chapter Cuttle lays out a total of seven distinct philosophical propositions, which he poses for consideration as approaches to museum lighting, some contradictory and some with the potential to overlap:

1. To make the artwork appear as it would have appeared to the artist at the time of its creation
2. To ensure that no damage due to light exposure will occur
3. To achieve the best possible appearance of the artwork
4. To provide optimum conditions for viewing art
5. To impart a sense of having seen ‘the real thing’
6. To assist viewers to understand the displayed objects and their reason for being there
7. For the lighting designer to establish a distinct and recognisable style

These propositions refer to museum lighting holistically, considering all aspects of museum lighting, but can be readily focused on the problem of colour

\(^{21}\)Credit for this phrase to Katherine Curran
2.3. Museum Lighting

appearance specifically. Before we narrow our gaze however, it is worth briefly considering a wide view of lighting attributes which may aid in the realisation of ‘good quality’ museum lighting. Consider Rosenfeld’s list of the five ‘controllable qualities’ in museum lighting; ‘intensity, movement [temporal artefacts], angle [modelling, avoiding glare and reflections], distribution [ambient lighting vs. spot lighting], color’ [261].

Whilst Cuttle’s propositions make for interesting discussions and enjoyable extended pondering, they are of limited assistance in the practical task of actually specifying lighting. Thankfully, a range of tools exist for the examination of the colour rendering properties of a light source, in the form of indices which aim to numerically describe an illumination’s effect on colour appearance of the objects of which it is tasked with illuminating.

Traditionally, colour rendering indices aim to offer a standardised method for calculating the colour differences induced by the substitution of a reference illuminant with a specific test source, and for comparing the relative merit of different test light sources on their ability to induce minimum change. In modern parlance this type of index should be referred to as a colour fidelity index, that is - one which is conceptually concerned with colorimetric reproduction. The term ‘colour rendering’ has come to encompass much more than just fidelity.

Diametrically opposed in some ways to fidelity are the indices which aim to quantify ‘preference’. In the simplest case, a preference index will aim to provide a value that is predictive of how an observer would rate a light source against other light sources.

Within and between these two groups there exist a range of different indices with subtly different aims and mechanisms for achieving these aims. For thorough overviews see Guo and Houser [132] and Houser et al. [150].

In practice, both of these philosophical approaches are mandated in current lighting guidance. For the most part, advice for lighting specification on the subject of colour rendering can be simplified to read ‘use lamps of above CRI 80 (referring tacitly to CIE $R_a$) but always test them visually before you buy in bulk’. Whilst this
may at first seem like sensible advice, upon further inspection it actually represents a serious contradiction, unless considered with heavy caveats. The problem rests in the fact that CIE $R_a$ is a fidelity index, whereas any visual inspection is likely to be performed by observing the appearance of an object under the test illumination without a reference. Fidelity aims to describe accuracy of reproduction, but this is a quality which is arguably not testable by visual inspection. This contradiction seems to perpetuate unnoticed in museum practice, with lighting specifiers often abstractly declaring to target a faithful/accurate/honest/impartial representation of objects, but practically choosing light sources based on visual inspection where preference is the only criteria. There is of course a range of approaches, ranging from pure reliance on indices to almost entire reliance on visual testing.

In conclusion, no one metric exists that would satisfy the divergent aims and philosophies of museum lighting. Several distinct types of colour rendering index exist, but the existing range fall into the broad categories of ‘fidelity’ or ‘preference’, with the latter being particularly poorly definable due to its variability in different environments, with different user functional requirements and different intrinsic preferences between different observers. Progress could be made by breaking these broad categories into smaller more manageable specific objectives.

## 2.4 Mathematical Methods

A small number of mathematical methods which may not be familiar to the reader are used within this thesis. They are outlined below.

### 2.4.1 K-means Clustering

K-means clustering is a method for cluster analysis, whereby unsorted data is sorted into $k$ distinct groups based on proximity to evolving anchors. Figure 2.31 shows the results of a k-means clustering on unsorted data, with colours indicating output cluster designations. It can be seen that the chosen clusters align well with what may have been chosen by a human observer. In this case the groups are relatively well separated, and so the task to the k-means algorithm was relatively undemanding.
Figure 2.31: A reproduction of Figure 7.13 from Page 281, which shows the results of a k-means clustering of the data from Figure 7.9 from Page 273 (which shows the actual groups.

The standard algorithm proceeds as follows:

1. $k$ random starting positions are assigned.

2. Each data-point is assigned a group based on which starting position is closest.

3. The mean location of the data-points in each group is computed.

4. Steps 2 and 3 are repeated, using the computed means instead of the original starting positions, until a stage where an iteration of these two steps results in no data-point changing groups. Once this occurs, the algorithm is said to have converged, and the resulting groupings are output.
2.4.2 Principal Component Analysis

A valuable primer on PCA in relation to colour technology is available from Tzeng and Berns [300].

Principal component analysis (PCA) is a dimensionality reduction method, used to reduce the number of variables within a dataset whilst retaining as much of the variance as possible, and often used to identify the correlated roots of variance.

![Figure 2.32](https://commons.wikimedia.org/wiki/File:GaussianScatterPCA.svg)

Along with similar techniques (such as single value decomposition), it is used extensively within the study of daylight SPDs [37, 145, 164, 212, 241, 243, 282] and
natural SRFs [1, 9, 58, 61, 89, 91, 98, 103, 135, 184, 209, 211, 212, 225, 245, 270, 312, 333].

In eras where the transmission of large datasets was troublesome, dimensionality reduction methods such as PCA held value as a method to summarise a dataset, with the understanding that a reader could reconstruct a pseudo-dataset with minimal data-loss from the provided principal components. One such example is the work of Judd et al. [164] where only the mean and first four characteristic vectors are provided\(^\text{22}\). The data from this study is replotted in Figure 2.33 and it can be seen that for their dataset this description does seem to offer a sensible summary of the data: the shape of the mean can be seen, and it can be seen that further variation principally occurs as a result of \(V_2\) which is a broad and relatively monotonous function, indicating that changes are likely to be a skewing of the spectral shape.

\[\text{Figure 2.33: The mean and first four characteristic vectors of Judd et al. [164].}\]

\(^{22}\text{This is not technically PCA but the related technique of Morris and Morrissey [226].}\)
2.5 **Interim Summary**

This chapter has laid out the most relevant developments in the research areas which the other chapters of this thesis build upon.

Chapter 3 builds upon our museum lighting knowledge by filling the gap in our understanding of how museum lighting is actually thought about and selected currently, and tries to identify the most fruitful avenue for future research which will allow the reduction of damage to objects in museums.

Chapters 4 and 5 target the interaction between colour constancy and ipRGCs, using some of the methodologies covered in the Section 2.1.4 and the knowledge of the physiology of ipRGCs outlined in Section 2.2 to build upon the experiments summarised in Section 2.2.3.

Chapter 6 investigates a new methodology (a variant of achromatic setting, described in Section 2.1.4) for performing colour constancy experiments, using museum spaces as test spaces due to their well controlled lighting environments.

Chapter 7 takes a more theoretical look at the relationship between colour constancy and ipRGCs using a computational methodology, and the mathematical methods described in Section 2.4.
Chapter 3

Interviews with Museum Professionals

The following work has been presented as an oral presentation at SEAHA 2016 [121], and published as a journal article: ‘How is museum lighting selected? An insight into current practice in UK museums’ [123]. A short summary of that article is presented here.

3.1 Introduction

In order to gain an understanding of how museum professionals currently specify lighting, interviews were conducted with 12 museum professionals representing 10 UK based museums, galleries or historic property management groups, in spring of 2016.

The goal of this work was to understand how lighting decisions were being made at this time, in order to increase our understanding of the decisions taken within practice, and also to identify gaps which would benefit from research development.

3.2 Methodology

The interviews were semi-structured around a set of questions composed by the author and their academic supervisors (reproduced in Garside et al. [123]). The choice to conduct these interviews in a semi-structured format as opposed to a
fully-structured one was based on the desire to allow unanticipated topics to enter the conversation, so as to limit the potential for important subjects to be neglected due to naivety of foresight. The conversational format of the interviews meant that the resulting data was qualitative rather than quantitative, which to an extent hinders comparison, but it was decided that the variety of job roles and institutional sizes, and the small sample size, already made quantitative comparison of limited meaningfulness for this investigation. Interviewees were granted anonymity, on the basis that this would allow them to answer questions more freely.

The interviewees were contacted through introductions from supervisors, personal connections, or cold emails. A small number represented UK-wide groups, but the majority represented London based institutions. Almost all held conservation based roles (the exception being an ‘exhibitions manager’) and noted that their principal responsibility regards lighting was controlling the safety of lighting regards its ability to cause damage to the objects held and displayed by their respective institutions. This generally involved monitoring and analysis of existing lighting systems and natural illumination, creating general guidance documents for the specification of lighting in their specific institutions and for loan items, and providing guidance and recommendations for the fitting out of new galleries or gallery refits. Few considered themselves directly responsible for the appearance of museum objects, considering this to be a creative decision outside of their remits. Future work considering the perspective of others involved in museum lighting, such as representatives from external lighting design companies, is required.

Whilst a small number considered themselves active in the area of lighting research, all interviewees were responsible for more conservation considerations aside from lighting, limiting the practicable level of specialisation. Many considered communication and dissemination a key part of their role, often noting they often found themselves in the position of needing to educate other teams within their institution on subjects including lighting.

“This is not my science, my job is pulling it out and presenting it to others.”
3.3 Summary of Responses

3.3.1 ‘Good lighting’

As asked what each considered to be ‘good lighting’, responses included ‘safe’, ‘invisible’ (you don’t notice it), and ‘lighting which is appropriate for your objects and your exhibition’ (noting the variability of requirements dependent on the particular object(s) being presented). Asked to present a list or range of priorities, many focused on the safety requirements of lighting. The principal safety concern for lighting was that it fell below specific illuminance criteria, dependent on the assumed sensitivity category of the object in question. The specific target values were generally those provided by Thomson [298] of 50lx for sensitive items and 200lx for less sensitive items (which are based on the visual preference work of Loe et al. [194]).

Considering a scale of other priorities, following the requirement for appropriate lux levels, considerations included: limiting/excluding UV, obtaining an acceptable CRI value, and time and capital costs associating with fitting and maintenance. Energy efficiency was also a driver, but this did not seem to be a factor within technology category, rather it was noted that many institutions are switching to LED because of increased efficiency, but that the difference in efficiency between one LED type and another was negligible compared to the saving in contrast to lighting technology which LED was replacing, which was often tungsten.

3.3.2 Roles and Tools

The range of roles played in the procurement of lighting varied amongst those interviewed. Whilst some created guidance documents which would then be passed on to estates teams and exhibition designers or lighting designers specifically, others had a much more hands-on role, testing specific lighting before installation or making recommendations on a case-by-case basis. It was noted that whilst relamping and retrofitting was generally handled by ‘in house’ teams, where new galleries or large temporary exhibitions were created it was common for an external lighting design company to be contracted to perform the work. When asked whether
recommendations were normally followed, the general reply was that recommendations for lux exposure and UV content were almost always followed, but other recommendations (such as for CRI or CCT) were more loosely interpreted. In many cases recommendations for CCT were not made.

When asked about the tools used to make such recommendations and choices, responses included references to guidelines and reference sources (most often Thomson [298], though also Druzik et al. [83] and British Standards Institution [32]), references to specific units such as ‘lux’ used in tandem with recommendations from guidelines such as above, indices such as the CIE general colour rendering index $R_a$ (referred ubiquitously to as ‘CRI’ by interviewees) and notes of specific conferences which were attended in order to stay up to date with the research on the subject. There was a general feeling that the current climate was one of swift technological change in lighting, which created an increased difficulty in staying up to date with developments. It was noted that attending conferences and industry workshops were very beneficial in assisting professionals to stay up to date.

"Things are moving so quickly that to rely on books which have taken two years to produce [does not suffice, because] things have moved on. Books (plus journals) used to be the main reference. Now things are moving at such a rapid pace."

3.3.3 Quantifying Safety

A range of techniques were used to qualify whether specific lighting was ‘safe’ or not. The most common practical technique was spot metering of lux values incident on specific objects, and selective dimming to drop incident lighting to the desired lux level in response to this. Some larger institutions with access to microfading equipment (whereby an intense light source is focused onto a very small area of an object, such that damage potential can be tested but in a manner which causes minimal damage to the object as a whole) were able to use this in the determination of sensitivity of specific objects. One interviewee provided details of a spectral power distribution based method for considering the safety specifically of phosphor based LED illumination, whereby the height of the blue peak was compared to the
height of the peak of the broader peak above 500nm, and if the former was more than three times the height of the latter, that lighting was singled out as potentially not safe. Other interviewees had heard of this criteria, and some used it as a rough guide, but one remarked that it was “fairly arbitrary”. One of the most succinct and perhaps astute responses was “what is ‘safe’?”. One interviewee referred to the website of Padfield [242] where considerations for the ‘RE%’ (‘relative spectral sensitivity normalised exposure values’), using the damage functions described by Aydinli et al. [10] are used.

The interviewees generally did not consider characteristics of the displayed objects such as geometry of illumination, broadly considering this to be the remit of a lighting designer or exhibition designer rather than a conservator.

3.3.4 Visual Testing

One of the most interesting, and perhaps surprising findings was the ubiquity of visual testing of lighting, generally performed prior to any large installation. For relamping, manufacturer supplied attribute values were generally relied upon as this required less time/effort and was cheaper. The most common procedure for new installations seemed to be for an informal and minimal visual test to occur before widespread installation. In a single case however, visual testing was actively avoided (on the basis that visual testing could not deliver meaningful insights where the aim was accurate rendering as opposed to visually pleasing rendering) and in another case large scale visual testing was performed, including many different types/brands of lamp and a large number of museum staff, and a final decision was made almost entirely on the results of this testing.

3.3.5 LED Usage

LEDs are now used, at least in part, in all the institutions involved in this survey. In several they are the primary lighting technology but in a small number they are used sparingly, only in applications such as the lighting of text information panels. In one they are used in a particularly minimal fashion, although this was attributed to the fact that the museum is moving location in the near future and thus capital
investments in building infrastructure were being avoided for the present time. There was no one specific brand or type which seemed to be ubiquitous across institutions, rather each institution appeared to have relationships with different manufacturers and suppliers.

Most interviewees were aware of warnings which had been issued and well publicised in the mainstream press [191] regards the potential of LED sources to be especially degrading for specific objects. Interviewees saw these warnings as controversial and likely unwarranted, and were confident that research had been conducted which cleared LEDs of causing an unacceptable level of damage in comparison with alternative technologies. When asked how they might assess a light source for safety, most replied that safety was assessed solely through use of an illuminance meter and lux targets, not through analysis of the SPD or any other lighting attribute. Those who did critically assess the SPD generally used no specific tools to do so, focusing attention on the wavelength of the spectral emission peak.

“*We never normally adjust the lighting type for a given artwork, we adjust the intensity.*”

“*For all lighting we measure the SPD, and check it is reasonable.*”

They key driver behind the adoption of LEDs appears to be energy efficiency increases and energy use reductions, as required by institution-wide directives, or as part of applications for planning permission. Secondary to this consideration, benefits noted include: decreased maintenance costs from extended lifetime of products, and a lack of availability of traditional bulbs, sometimes due to specific legislation which has in effect phased out some traditional technologies. One element holding back some interviewees from further investment was a residual feeling that this new technology was not yet fully proven. Many pointed out that the claims made regards extended lifetime of LEDs were yet to be proven in real world environments due to the relatively new nature of the technology. Some also noted the high costs associated with having to change the underlying lighting
infrastructure, where retrofitting wasn’t possible or appropriate. A final note on this section - some interviewees were unsure about the ability of LEDs to remain colour stable over the long expected lifetime of the products.

Interviewees reported that visitors had not generally responded to any changes in lighting technology, and this was taken to mean that any switch to LED had not been negatively received. It was inconclusive whether or not the technology was positively received however. This could be a meaningful avenue for future work.

In terms of the professionals own opinions of LED lighting, all seemed favourable, though it was unclear how much of this effect was caused by extraneous or related effects such as a placebo effect due to the excitement of new technology, or different chromaticities or brightnesses of replacement technologies.

“I like what I’ve seen. The galleries where we have just LED spots, I feel happier. I went to [another institution, with abundant LED lighting], I really like the galleries where they had LED lighting, and it was more of a gut feeling rather than something which I could put my finger on, but actually, it felt cleaner to me.”

3.3.6 Spectral Tuning of Light Sources

Whilst all interviewees were interested in the idea of making objects appear brighter whilst reducing the level of damage caused, the point was made that whilst spectral tuning might benefit a prototypical object, it will not necessarily benefit real objects in real environments. It was also noted that the use of lux in making decisions was particularly problematic here, where varying the location of a blue peak could easily increase the level of damage but reduce the lux value. Most interviewees had a basic understanding of colour temperature, and a limited understanding of chromatic adaptation. The colour temperature of lighting was generally not seen as a conservation issue (though there were exceptions to this), and rather as a creative consideration. One interviewee noted that it was manipulated to great effect by external lighting designers in order to create specific effects or atmosphere.

The justification for CCT specification values generally appears to stem from two routes. Firstly, from guidance documents such as those provided by The Getty
and the Canadian Conservation Institute [83], and secondly from a desire to match existing lighting, either daylight or more commonly tungsten (at around 3000K). It was rarely considered as a means to control damage, or as a way to enhance visual appearance, by those interviewed. One interviewee referred to the work of Kruithof [180] and Scuello et al. [266, 267] as justification for choosing low CCT illumination similar to tungsten. No specific issues relating to colour temperature were raised by interviewees.

3.3.7 Colour Rendering

The interviewees were very interested in the subject of accurate representation, and almost all seemed to regard accurate object representation as a key priority. The figures for $R_a$ quoted in internal documents at each institution were either 80, 85 or 90, as a minimum figure. In one institution, an $R_w$ value was calculated for each proposed light source (the R value of the test colour sample with the colour shift of greatest magnitude, aka ‘worst’) and a cut-off of 80 imposed. However, most interviewees seemed unsure of the practical relevance of $R_a$, with many considering it a rough guideline which would be considered secondarily to a visual inspection of lighting.

Those who were particularly interested in the subject gave the impression that whilst the subject was considered philosophically in great detail, the tool which was actually used to analyse the colour rendering of a light source was still generally just $R_a$. A few interviewees were aware of TM-30-15 [158], and whilst it was respected, it was questioned whether it represented a real improvement over $R_a$.

On the subject of lighting philosophy, the opinions encountered generally aligned with the mechanics of $R_a$. That is to say; given a choice between illuminating an object such that it was beautified, visually restored (to a previous condition) or simply presented as it would appear under daylight/tungsten illumination, most opted for the final option. Whilst there was clear interest in the other options, and other creative ways in which to consider colour rendering, it was generally believed that the role of the museum should be to represent objects in an un-biased
manner, and thus a fidelity index was an appropriate tool for discussing a light source’s colour rendering properties. In the case where specialist lighting was used to special effect, the opinion was noted that “you have to be very clear about what you are doing and why” in order to maintain the reputation of the museum as an arena for honest and unbiased representation.

3.4 Conclusion

Generally, the interviewees believed that visitor requirements were being met (although there is often difficulty in defining exactly what visitor requirements actually are) and no specific tool or technology was proposed that would provide a clear benefit. Several interviewees mentioned that a way to improve the accessibility of colour rendering indices (in terms of the ease with which they could be understood and applied) would be appreciated. No interviewees knew of recent surveys similar to the present one.

3.5 Interim Summary

The impact that these interviews had on the research which followed is thus; it became clear that CCT was a tool which could be used to reduce damage, but which was not being used at the time. One of the barriers to use was a lack of understanding of how CCT interacted with other visual appearance properties and preference. It therefore seemed valuable to attempt to extend our understanding of chromatic adaptation and colour constancy, with the hope that this would allow museum professionals to limit damage through specification of lower values of CCT.

The following chapters all seek to extend our knowledge of colour constancy and chromatic adaptation.
Chapter 4

Large Sphere Experiment

The work presented here has been presented as an oral presentation at AIC 2013 [205] prior to the author’s involvement, and as an oral presentation at ICVS 2017 [120] by the author.

Code and data is provided: https://github.com/da5nsy/LargeSphere

4.1 Summary

The goal of this experimental work was to examine the effect of different wavelengths of light upon chromatic adaptation. Our hypothesis was that ipRGC stimulation may need to be considered in order to fully model the induced adaptation, with the null hypothesis being that chromatic adaptation can be fully accounted for by cone and rod mechanisms. If evidence of a melanopic input to chromatic adaptation was found, it may help to explain conflicting results in previous experiments which sought a ‘preferred CCT’, which may in turn allow for control of CCT in museums to be used more extensively as a means to control damage to objects.

This experiment is of a similar type to those discussed in Section 2.1.4.3. Within a Ganzfeld viewing environment, illuminated by one of 16 different wavelengths of near-monochromatic light, observers performed an achromatic setting task, controlling the chromaticity of a display visible in the central field through a 4° circular aperture with two handheld sliders. Under these conditions it would be expected that an observer’s chosen achromatic point would correspond in hue to the adapting field, and be of a saturation somewhere between a nominal objective white point
and the adapting stimulus. If melanopsin were involved in chromatic adaptation we may expect unusual results for the part of the spectrum that melanopsin is most sensitive to (roughly 480nm).

Two different analyses were performed, with neither providing comprehensive support for rejecting the null hypothesis. However, it is noted that several assumptions are implicit in the experimental design, and that the experiment samples only a small region of the potential search space for melanopic input to adaptation (See Section 4.4).

This project was designed before the author arrived at University College London (UCL), and data from two participants had already been collected. Data collection required at least 16 hours commitment from observers, and so the only observers up to that point had been LM (one of the author’s academic supervisors), who initiated the experiment, and TR who was a student in the Medical Physics department. The original goal for my involvement in this project was that I should be a third observer, and assist in the data analysis. However, following the collection and initial data analysis of my own data, it appeared that there had been a technical fault during this run of data collection, and my data was deemed corrupted. Thus, my only contribution to this work is an extension to the data analysis started by LM, upon which I shall focus on in this chapter. The issue does not seem to have affected the data from the other two observers.

4.2 Methodology

4.2.1 Hardware

A hollow fibreglass sphere of approximately 750mm diameter was prepared with three holes; the first for an observer’s face, the second (above the observer’s head) for a light source to illuminate the Ganzfeld, and the third (opposite the first) through which a small portion of an LCD screen could be seen (A Fujitsu-Siemens SCALEOVIEW D19-1, SN:YE5L006100). A schematic is shown in Figure 4.1.

The screen was characterised such that calibrated stimuli could be generated. This was done by taking spectral measurements at 21 levels of pixel value for each
channel, and later interpolating to create a full look-up-table to transform from pixel value to XYZ values and vice versa.

The interior of the sphere was painted with RAL 7040 dulux vinyl matt grey, of approximately 38% reflectance.

Illumination was provided by a Kodak slide projector with a tungsten-halogen light source, filtered through one of 16 near-monochromatic filters, ranging in 20nm intervals from 400-700nm inclusive (Figure 4.2). During each session, only one filter was used and the surround illumination remained the same throughout. Measurements of the internal illumination were taken with a Photo Research PR-650 (PR650) device, plotted in Figure 4.2. The illumination has been described further in MacDonald and Roque [205]: “The average luminance of the surrounding chromatic adapting field ranged from a maximum of 0.75 cd/m² at 560 nm to less than 0.05 cd/m² at the ends of the spectrum, corresponding to a retinal illuminance through a pupil of diameter 8mm ranging from 38 trolands (max) to less than 2.5 trolands, meaning that the viewing environment was in the upper mesopic range”. No effort was made to ensure that light from inside the sphere did not hit and/or

Figure 4.1: The hardware design. Illustration courtesy of Lindsay MacDonald.
reflect from the LCD display.

![Graph](image_url)

**Figure 4.2:** The illuminations within the sphere, created by filtering light from a slide projector, measured as reflecting from a point just to the right of the aperture through which an observer would view the screen. Measurements made by Lindsay MacDonald.

### 4.2.2 Observer task

The observer sat on one side of the sphere with their face inside the sphere (as shown in Figure 4.3), such that nothing outside of the sphere was visible. On view on the opposite side of the sphere was a circular 4° aperture onto an LCD screen, upon which a random colour was visible (See section 4.2.3 for further details of the randomisation starting routine). Surrounding the aperture, the rest of the ganzfeld was illuminated by light from the slide projector, filtered through one of 16 near-monochromatic filters (Figure 4.2). It was the observer’s task to use two handheld sliders, which controlled the chromaticity of the screen, to make the appearance of the screen achromatic (an achromatic setting task).

On average it took observers roughly 20 seconds to make a selection. Once
the observer was happy with the achromacy of the patch, a button was pressed to record the setting and a new random colour would be presented. The first displayed colour was at CIE L* (of CIELAB and CIELUV) of 85, with subsequent colours descending by 5 L* until 10 L*.

This sequence was repeated 10 times per session. Per session observers made 10 selections at 16 lightness levels (160 total). Observers performed 16 sessions (2560 selections total), one session for each surround adapting wavelength. The overall protocol is visualised in Figure 4.4. Observers found sessions quite fatiguing and generally did not wish to do more than two or three sessions per day. A brief break was generally taken between sessions, though no minimum time for such was prescribed.

For one observer, in an additional (17th) session the narrow-band filter was replaced by a neutral density filter, to produce an achromatic adapting field.
4.2.3 Stimulus specification

The stimulus was controlled via a MATLAB script, which read the input of two sliders and a button via a ‘Phidget’ interface. The two linear sliders provided values of between 0 and 1000, and these values were converted to approximate CIELAB co-ordinates in such a manner that the slider maxima corresponded to the Natural Colour System (NCS) unique hue positions as computed by Derefeldt and Sahlin [78]. In this manner the sliders could be considered as moving along opponent axes between red and green (via the CIELAB origin), and blue and yellow (via the CIELAB origin) respectively.

These values were transformed into XYZ values, with white references of \( \text{[XYZ} = 99.04, 100, 151.30] \) for observer LM and \( \text{[XYZ} = 94.97, 100, 98.15] \) for observer TR. The white reference for LM related to a screen characterisation performed around the time that the initial measurements were made. It is assumed that the same is true of the white point used for observer TR, although this characterisation data is no longer available. These values were then converted to sRGB values, and output to screen.

The generation of random starting colours was achieved by modulating the nominal zero-point on each slider scale, where the default zero point is considered as 500, sampling from a uniform distribution between 250 and 750 for each presentation. The slider position was then considered relative to this new zero-point. The effect of this can be considered as such: if the observer were to leave the slider in

<table>
<thead>
<tr>
<th></th>
<th>Repeat at 16 different lightness levels</th>
<th>Repeat 10 times (in order to see progression over time)</th>
<th>Repeat 16 times with different surrounds (max 4 per day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roughly 20 seconds</td>
<td>Roughly 6 minutes</td>
<td>Roughly 1 hour</td>
<td>Roughly 16 hours</td>
</tr>
</tbody>
</table>

Figure 4.4: The experimental protocol.
4.2. Methodology

a central position throughout (make no selections) they would be provided with random colours drawn uniformly from within bounds either side of the objective white-point. Since, hopefully, the observer was making selections, each new ‘random’ colour would be biased towards the previous selection (based on where the sliders had been left from the previous selection). It should be noted that this new colour was not entirely independent of the previous selection, due to this bias.

4.2.4 Data Processing

Data were calibrated in the following manner: the recorded RGB values of the observers’ selections were bounded (values above 1 or below 0, which occurred when observers made selections which were outside of the sRGB gamut, were brought within range, with an ‘absolute’ rendering intent), quantized to 8-bits, and converted via look-up-table to CIE 1931 XYZ tristimulus values. From these, xy chromaticities and CIELAB values were computed (with the white point of the display, as loaded from the characterisation file, used as the white reference).

A set of data referred to as the ‘baseline’ data was generated, where there was no observer input, to consider the range of possible responses that an observer could make. The element of code which provided new random starting positions was excluded for these sessions. This data was processed in exactly the same manner as the observer data, and is shown in Figure 4.5, where it can be seen that the chromatic gamut increases as \( L^* \) increases (due to a factor \( c f \) in the display code which aimed to scale chromatic space with \( L^* \), mirroring the shape of CIELAB). It can also be seen that the gamut boundary is sometimes reached at higher levels of \( L^* \), with some of the vectors curving to remain within gamut.

Two distinct approaches were taken to data analysis. The first attempted to process the data in a chromatic space, with the reasoning that under the null condition chromatic selections should simply correspond to the chromaticity of the surround adapting illuminations, presumably with some sort of gain function applied. If it could be shown that this relationship was not as expected, in a manner which might suggest involvement by other mechanisms (meaning rods or ipRGCs), then this could be considered as evidence against the null hypothesis.
Figure 4.5: The baseline dataset. This data represents the condition where there is no observer input. The points on the curves are the different values of L* presented during a session. Each curve represents a single ‘session’. The two sliders are left at their maximum (1000), minimum (0), or neutral (500) positions (9 combinations). This was done computationally; no actual slider input was used. In the legend the first number denotes the slider A position, and the second denotes the slider B position.
4.3 Results

4.3.1 Summary

Visually summarising the results for even a single observer is difficult due to the high number of variables and collected datapoints. In the following plots data is averaged over the 10 runs within each session (under each adapting wavelength). In Figures 4.6 - 4.8 data from LM, TR and the author (DG) is presented.

It can be seen in the data for all observers that there is a clear $L^*$-dependent shift. Interestingly, the precise nature of this seems different for each observer; the data of LM clusters tightly around [0,0] for low values of $L^*$ and then generally moves north-west as $L^*$ increases, before returning towards the origin at the highest values of $L^*$. For TR the shift is monotonic and roughly south-west.

Within this pattern it can be seen for observers LM and TR that there is a causal relationship between the adapting wavelength and the selected chromaticity. This is most easily seen in the top right plots of Figures 4.6 and 4.7. In both cases a rough circle can be seen for both of the values of $L^*$ shown.

In Figure 4.8 it can be seen for observer DG in the lower-right plot that the data splits into two distinct groups, one of which is considerably lower in $b^*$ than is seen for either of the previous observers, or would generally be expected. It is suspected that there was a screen calibration issue which led to offsetting of the screen output in an unpredictable session-by-session fashion. It was found that a basic offsetting applied selectively to those sessions which appeared to be affected could ‘correct’ for this issue, but without a better understanding of the issue this dataset is considered unsuitable for further analysis.
Figure 4.6: The dataset of observer LM. In all plots, the average CIELAB values are taken over the 10 repeats within each session. Top left: an overview of the chromaticity of selected points. Connections between points indicate that they are from the same run (under the same adapting wavelength). Colouring of points and lines indicates the adapting wavelength, though whilst there is an approximate correspondence between the colour of the line and the appearance of the adapting field this is only meant as a means to differentiate between the different lines and is not in any way an accurate representation of appearance. Top right: as for top left but only for points recorded at specific values of L*. Colours are as for top left, but lines linking points now represent points of like L*. This plot is included to show the relationship across wavelength of the adapting field. Lower left and right: Other perspectives upon the CIELAB projection. Data as for top left.
4.3. Results

4.3.2 Variability over time/repeats

Not represented in the above plots is the way in which responses varied over time within each session, averaged over \( L^* \) (the previous plots were averages over time). Figures 4.9 and 4.10 show the calibrated CIELAB values for observers LM and TR.

\( L^* \) should remain steady throughout (this was set), with minor differences introduced presumably due to differences between the screen and sRGB, 8-bit quantization, and any selections where a gamut boundary was reached. Both \( a^* \) and \( b^* \) follow the broad trends which would be expected given previous figures. Newly visible in these figures is the manner in which responses change over time. For both observers LM and TR the first two or three repeats seem to be distinct from the rest of the set, suggesting that adaptation had not yet reached a steady-state.

Figure 4.7: As per Figure 4.6, but with the data of observer TR.
Figure 4.8: As per Figure 4.6, but with the data of observer DG.

during this time. Further quantitative analysis is provided in the following section.
Figure 4.9: CIELAB co-ordinates across time (repeat number) for observer LM. Top plot is $L^*$, middle $a^*$ and lower $b^*$. 
Figure 4.10: As per Figure 4.9, but with the data of observer TR.
4.3.3 Chromaticity-based analysis

The CIELAB co-ordinates for the adapting fields were computed from the measurements shown in Figure 4.2, relative to the white point of the display for observer TR, and are presented in Figure 4.11.

![Figure 4.11: The CIELAB values for the surround adapting illuminations, calculated from the measurements shown in Figure 4.2, taking the white point of the screen (for the TR trials) as the white point.](image)

If the surround fully controlled adaptation, and observers were fully adapted, it would be expected that observers would select the chromaticity of the surround as their neutral chromaticity. It is unlikely that either of these statements is correct, but we would still expect selections to be impacted by the chromaticity of the surrounds to some extent. If we re-plot the selected L* values from the top-right sub-figures of Figures 4.6 and 4.7 atop the data of Figure 4.11, we can visualise the correspondence between the observer selections and the surrounds.

\[1\] Note that the data of Figure 4.11 is transformed for Figure 4.12. The same measurements data was used for the surrounds, but the normalisation factor was different since the white point of the display was used when calculating CIELAB co-ordinates.
Figure 4.12: The CIELAB co-ordinates for the surround illuminations, relative to the white point used in the LM trials (black line), and the data from the top right sub-figure of Figure 4.6, showing the CIELAB co-ordinates of the observer selections for 20 L* and 60 L* (dashed and dotted grey lines respectively).

For both observers we see that the pattern of responses corresponds very well to the pattern of the chromaticities of the surrounds. For both observers there is a scaling and offset but considering the L* dependent shifts seen previously, and the way in which isoluminant planes through CIELAB change shape with changing values of L* this is to be expected.

Whilst there is some marked variation from a perfect replication of the surround CIELAB co-ordinates - for example the 500nm and 600nm points for LM, and the 520nm and 540nm points for TR, these variations seem to be in line with the level of noise in the responses, and attribution to a specific cause (such as melanopsin or rod input) cannot easily be achieved, since many variables are confounded.
4.3.3.1 Colour Constancy Indices

Colour Constancy Indices (CCIs) can be calculated by comparing the distance between a pre-adaptation point (generally some sort of objective white point) and the post-adaptation point (the participant-selected achromatic point, or some average thereof), with the distance between the pre-adaptation point and the nominal ‘ideal’ match (which in this case would be the chromaticity of the surround adapting field). The CCI is calculated as:

\[
CCI = 1 - \frac{b}{a} \tag{4.1}
\]

where \(b\) is the distance between the post-adaptation point and the ideal match, and \(a\) is the distance between the pre-adaptation point and the ideal match\(^2\).

There are multiple reasonable options for which value to use as a ‘pre-

\(^2\)For further discussion see Foster [113, Section 4.1, pg. 681].
adaptation point’. First, the origin of the space within which selections are made (different for each observer) seems to be a possible option; this corresponds to the central point on each slider over time for each individual. However, though the set-up ascribes some value to this point, it is not definitively linked to the settings that observers made; it can be seen in Figures 4.12 and 4.13 that there seems to be no particular relevance of the point [0,0]. A second option would be to use the measurements made under a neutral density filter. However, again there is no actual significance of these values - a neutral density filter could be slightly chromatic and still be labelled as a neutral density filter, and even if it were perfectly spectrally neutral, its designation as a gold standard ‘neutral’ only actually passes on responsibility to the chromaticity of the projector lamp, which is under no obligation to be especially ‘neutral’. The third option is to use the average setting value, which has no specific logical background, but is vastly more practically relevant than the previous two options. This third option was chosen for future analyses.

Averaging over time for each observer, and using the average response for each observer as the pre-adaptation point yields CCIs as shown in Figures 4.14 and 4.15. Only data for L* of 20 and 60 is plotted for clarity, in keeping with previous figures. It can be seen that there are common trends across wavelength at the different values of L*.

It is highly unusual for values of CCI to be below 0; this indicates that the selected post-adaptation point is further from the pre-adaptation point than the ideal match. Normally, assuming that adaptation occurred on the same vector as that connecting the neutral point and the ideal match, this would mean that the observer had over-adapted, something which is very unusual. Additionally, results like this generally aren’t seen because observers are adapted to highly saturated adapters, often on/near the spectral locus, and colours outside of this simply don’t exist to be chosen (in a linear space).

We see such results here for a number of reasons. Firstly, we are not in a linear space. In CIELAB the chromatic gamut increases as a function of L*, meaning that a high L* value can be outside of the gamut of a set of low L* primaries. Secondly, in
4.3. Results

concert with this non-linearity, the slider ranges were fixed to represent a broader range of a* and b* values at higher L* values (otherwise it would have felt as though a specific movement at a low value of L* would have resulted in a much greater chromatic shift than that same movement would have done for higher values of L*). The effect of this is visualised in Figure 4.5.

Additionally, it appears as though there is substantial offsetting and L* dependent shifting, which call into question the appropriateness of such a metric. It should also be noted that in this current analysis, averages are taken over time, which obscures and adopts any underlying trends which time may influence. It seems as though there is a risk of obscuring more than is revealed through use of such a metric.

However, it is curious to see a dip in the results for both observers at 500nm, which is roughly where we might expect to see an effect should there be an effect of melanopsin (which theoretically peaks at around 480nm, but is predicted to
have an increased value of peak sensitivity as a function of pre-receptoral filtering). Based on Figures 4.12 and 4.13 this was not anticipated. However, without a clear prediction for what effect we would expect melanopsin to have (in terms of the direction or magnitude of effect) I suggest caution in interpreting this as evidence of an effect. This peak could also be the result of rod-based intrusions (the peak of the rod SSF is 507nm). It is unclear what magnitude and vector of effect should be expected from rod intrusion.

Averaging over wavelength and time allows us to visualise the effect of $L^*$. Here, instead of calculating the CCI, a simpler measure is used: the distance from the pre-adaptation point (the average of all recorded achromatic points, per observer) to the post-adaptation point. This gives us a more direct impression of the extent of adaptation, without the assumption of adaptation vector angle. In the context of Equation 4.1 this value could be denoted $c$, as it represents the final side of the triangle $abc$. 

**Figure 4.15:** As per Figure 4.14 but for observer TR.
4.3. Results

It is assumed, based on the analysis presented in Figure 4.5, that as \(L^*\) increases, the length of these vectors shall increase, simply as a result of the experimental set-up. This is shown to be the case in Figures 4.16 and 4.17, with near monotonic increases as \(L^*\) increases for both observers.

![Graph showing average distance (in the a*b* plane of CIELAB) between pre-adaptation point and post-adaptation point for different values of \(L^*\) for observer LM. Error bars are standard deviation.](image)

**Figure 4.16:** Average distance (in the a*b* plane of CIELAB) between pre-adaptation point and post-adaptation point for different values of \(L^*\) for observer LM. Error bars are standard deviation.

Averaging over wavelength and \(L^*\) allows us to visualise adaptation over time. Again, we would assume that as time increases (technically, we use repeat number here as a rough surrogate of time) we should expect to see an increase in the vector distance between pre-adapt and post-adapt. It is possible that a hint of this trend is visible in the data presented in Figures 4.18 and 4.19 but the level of noise is very high as can be seen from the measures of standard deviation.

A three-way ANOVA performed upon the data, treating wavelength, time and \(L^*\) as independent categorical variables found a significant effect of each, as shown in Figures 4.20 and 4.21, with a level of \(\alpha\) of 0.05.

Whilst variables were treated as categorical in the above analysis, there would be an argument for treating each as a continuous variable. However, several fac-
Figure 4.17: As per Figure 4.16 but for observer TR.

Figure 4.18: Average distance (in the a’b’ plane of CIELAB) between pre-adaptation point and post-adaptation point for different repeat numbers for observer LM. Error bars are standard deviation.
4.3. Results

Figure 4.19: As per Figure 4.18 but for observer TR.

Figure 4.20: The multi-way ANOVA output table for the LM data, where X1 is wavelength, X2 is repeat number (time), and X3 is L*.

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum Sq.</th>
<th>d.f.</th>
<th>Mean Sq.</th>
<th>F</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>162069.4</td>
<td>15</td>
<td>1017.16</td>
<td>69.36</td>
<td>4.05067e-174</td>
</tr>
<tr>
<td>X2</td>
<td>616.7</td>
<td>9</td>
<td>68.52</td>
<td>4.6</td>
<td>4.74001e-36</td>
</tr>
<tr>
<td>X3</td>
<td>14127.7</td>
<td>15</td>
<td>941.85</td>
<td>68.24</td>
<td>2.58106e-162</td>
</tr>
<tr>
<td>Error</td>
<td>37532.6</td>
<td>2520</td>
<td>14.69</td>
<td>14.69</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>67546.4</td>
<td>2559</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.21: As per Figure 4.20 but for observer TR.
tors would need to be considered. Foremost, whilst wavelength is nominally a continuous variable, in this experiment each wavelength category had a different level of radiance, which means that caution should be taken in assuming their equivalence. It is also possible that each filter might have a meaningfully different spectral transmission profile, specifically the band-pass width (see Figure 4.2).

Caution is also required regarding the assumption of independence of measurements. Due to the nature of chromatic adaptation, it would not be possible to interleave conditions, and so wavelength is further confounded with various other factors; date, time of day, and all manner of secondary factors relating to these (whether the observer has eaten recently for example).

It would be of interest to assess whether the contrast between the surround and the selection area influenced settings, but for this specific experiment the contrast is confounded by wavelength and L* and there does not seem to be a clear way to examine the influence of contrast directly.

4.3.4 Spectrum-based analysis

This analysis aimed to leverage the fact that we have access to estimates of the spectral emission of the screen for each RGB value, and thus can calculate the relative cone/rod/ipRGC catches for each achromatic setting. This in turn allows us to test to what extent the results seen can be explained simply by scaling cone mechanisms (a diagonal, or Von-Kries-type transformation) and to ask whether adding additional inputs to the model (rods and/or ipRGCs) improves our ability to explain the measured results.

The first stage of this analysis was to generate simulated data which represented the situation whereby there was only simple Von Kries adaptation. Under this situation the estimated cone catches for the achromatic matches would be equal to the SSFs of the cones, linearly scaled by the radiance levels of each surround adapting field. If the radiance levels were equal at each wavelength interval, the simulated data would be equal to the cone SSFs. If, hypothetically, one wavelength interval were vastly higher in radiance, we would expect a correspondingly higher adaptive effect, which would result in a higher level of activation required in order
for an achromatic visual appearance. In this way, we predict what results may look like if the only type of adaptation occurring was a simple Von Kries / diagonal scaling.

Practically, this is accomplished by element-wise multiplication of each sensor SSF by the measured emission from each adapting surround, as per Equation 4.2. The absolute scaling, and the relative inter-sensor scaling, is irrelevant due to the freedom that will be allowed later in the analysis.

\[ s_i = p_i \odot e \] (4.2)

where \( s \) is the simulated required sensor catch for achromacy (with the index \( i \) denoting the sensor), \( p \) is the sensor SSF, and \( e \) is the measured emission from each adapting surround. The CIE 2006 10° observer fundamentals were used, and the results are visualised in Figure 4.22.

A comparison was then made between this data and a set of real data (Obs = TR, averaged over time (entire run, no exclusions), averaged over \( L^* = 35:60 \)). This real data had been transformed from the recorded RGB values of achromatic matches into LMS values. It can be seen that there is a considerable difference between the simulated data and the real data (Figure 4.23). S-cone data shows the closest match, with the predicted peak at 460nm\(^3\) being mirrored in the real data. This peak appears to bleed into the (real) M-cone data, and the simulated data for L and M-cone data shows very little correspondence to the collected data. Correlation coefficients between the simulated and this specific real dataset are 0.1399, -0.2509, 0.3164 for L, M and S respectively.

In order to understand the way in which adaptation may be crossing between channels, or the way in which we may have not properly isolated our channels (it is unclear exactly how much freedom an observer truly has to move around the response space) a brute-force method was used to find combinations of the above simulated data which would best fit the real data.

\(^3\)Note that this is not at the peak sensitivity of s-cones (which would appear at 440nm for this dataset at 20nm intervals) but rather at the peak of the s-cone sensitivity function multiplied by the SPD, as plotted in Figure 4.22.
**Figure 4.22:** Simulated data for a basic Von Kries observer. In this figure white represents a high response. For example, with 560nm peripheral stimulation we would expect an observer to pick a colour with high L-cone activation as achromatic, assuming that the sensitivity of L-cones had been suppressed and thus higher activation was required to reach a neutral point (see peak roughly in the centre of the top bar).

10000 random sets of weighting values (30000 values total) between -25 and 25 were generated. These weightings were applied to the simulated responses and the results were additively combined. The correlation between this new random combination and each channel of the real data was computed. The top performing randomly generated combinations were selected and are presented in Figure 4.24. These particular combinations were created through cross-combining the original simulated data (top row of Figure 4.23), in the ratios shown in Table 4.1 and correlated with the real data to extent of the following coefficients: 0.9126, 0.8861, and 0.7726 for L, M and S respectively. These are much improved over the

---

4 An analysis showed that the absolute range of these figures was unimportant, since we were looking for correlation with the real data rather than absolute correspondence. Thus they are listed here only to assist the reader in understanding graphs such as Figure 4.25.

5 \((X \text{ amount of simulated L}) + (Y \text{ amount of simulated M}) + (Z \text{ amount of simulated S})\)
4.3. Results

Figure 4.23: A comparison of simulated data (top row) and real data (bottom row). Data from the top row is as per the top three bars of Figure 4.22.

Table 4.1: Optimal weights to fit the specific real dataset used.

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>M</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>18.4069</td>
<td>-23.2578</td>
<td>-10.9817</td>
</tr>
<tr>
<td>M</td>
<td>-13.0477</td>
<td>9.8327</td>
<td>10.6844</td>
</tr>
<tr>
<td>S</td>
<td>-2.7633</td>
<td>-17.6798</td>
<td>8.3036</td>
</tr>
</tbody>
</table>

Example: Image in top left of Figure 4.24 (L) was created by combining $18.4069 \times$ the original simulated L (Top left of Figure 4.23), $-23.2578 \times$ the original simulated M and $-10.9817 \times$ the original simulated S.

The top performing 0.2% of the randomly generated combinations are presented in terms of their components (analogous to plotting the values in 4.1) in Figure 4.25, and the entirety of the results for the randomised sampling presented in Figure 4.26.

It can be seen that to reconstruct the real data for L, a high amount of simulated L, and a low amount of both simulated M and S are required, though from Figure
Figure 4.24: A comparison of randomly generated combinations (top row) whereby channels were freely mixed from basic Von Kries simulated data (top row of Figure 4.23) to best correlate with real data, and real data (bottom row) (repeated from Figure 4.23).

4.26 it can be seen that the requirement for low S is less stringent. It can also be seen that the amount of L required seems related to the amount of M required (from the way in which the lines cross at a point).

A similar but opposite trend is visible for M.

For S, there is a narrow range of successful values for L (negative but close to 0), a larger range of strongly negative values for M, and a range of positive values for S. The reciprocal relationship between L and M seen in the reconstructions of L and M is no longer visible, but instead there is a new reciprocal relationship visible between M and S, though examination of Figure 4.26 suggests that this is not as important as in the case of L and M.

It is reassuring that in each case, successful random combinations used high positive levels of the target signal. For both L and S the target signal was the only positive weighting, with M taking positive weights of both M and S.
4.3. Results

Figure 4.25: Top 0.2% performing randomly generated combinations, presented in terms of the weights of the original simulated data that they use. The subfigure on the left represents the weights needed to reconstruct the real data for L, the middle - M, and the right - S. Colour coded such that dark blue is the highest performing and yellow is the worst performing (of this highly performing subset).

4.3.4.1 Adding rods and ipRGCs

In order to investigate whether rods or ipRGCs were playing a role in adaptation as measured by this dataset, the analysis was re-run with the additional rod input, additional ipRGC input, and both rods and ipRGCs as additional inputs. See Figure 4.22 for a visualisation of these channels. The results of re-running the computations following the addition of these inputs is shown in Table 4.2. Minor increases in correlation are exhibited. However, it should be noted that one would expect to see at least a minor increase in performance from practically any additional signal, so long as it was independent from the already accessible signals.

Further, it is not clear to what extent the gains exhibited in Table 4.2 are due to noise within the computations; the randomly generated combinations are set
Figure 4.26: As per Figure 4.25 but for all randomly generated combinations. The trends seen in Figure 4.25 are visible, and the range of these trends (extending into the poorer performing randomly generated combinations) can be seen. The colours are rescaled such that yellow now represents the worst performing randomly generated combinations of the entire set. Plots are plotted ordered by success and so lines representing successful randomly generated combinations will overlay lines representing poorer performing ones.

<table>
<thead>
<tr>
<th></th>
<th>Just cones:</th>
<th>+ rods:</th>
<th>+ ipRGCs:</th>
<th>+ rods + ipRGCs:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.9126</td>
<td>0.8861</td>
<td>0.7726</td>
<td>0.9156</td>
</tr>
<tr>
<td></td>
<td>0.8931</td>
<td>0.8881</td>
<td>0.8055</td>
<td>0.9326</td>
</tr>
<tr>
<td></td>
<td>0.9218</td>
<td>0.8867</td>
<td>0.8100</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: Correlation coefficients for various conditions incorporating additional signals.
4.3. Results

via a random number generator which is re-set each time the script is run for reproducibility, however there is nothing to stop the randomly generated values for the additional input runs performing better purely by chance alone. In order to investigate this, the above extensions were re-run 100 times each, and the top performance for each skimmed and saved. The results of this are presented in Figure 4.27. From this figure it can be seen that there is a real and clear benefit from the inclusion of the additional signals, and from inclusion of both of the additional signals.

![Figure 4.27](image)

**Figure 4.27:** Raincloud plot [5] showing the results of re-running the extended analyses 100 times and skimming the best performer from each. Red points and probability density function represent ‘just cones’, green - ‘+ rods’, blue - ‘+ ipRGCS’, black - ‘+ rods + ipRGCS’.

4.3.4.2 Spectrum-based Analysis Discussion

There is a clear distinction between the simple simulated response functions, and the recorded data. It is interesting that a simple linear recombination can improve the correlation so greatly. It should be remembered however that this type of
post-hoc fitting is liable to delivering whatever results a researcher might hope to find.

Taken at face value, the results suggest that the principal drivers of adaptation are not at the cone level, but rather at a higher level, once cone inputs have been combined. The results mirror what might be expected of these higher level signals - there appears to be a single signal for L and M, roughly mirrored between the two, with a reciprocal trade-off possible between L and M for both, and the S cone signal takes positive weights for S, and negative weights for both L and M, but with a curious hint of a reciprocal trade-off between S and M.

However, it is unclear to what extent these relationships may arise due to limitations of the experimental set-up; it is possible that a rise in one signal is yoked through hardware limitation to the rise or fall in another. Future investigators should consider whether this effect is modellable.

Though there is a demonstrated ability of additional signals to improve the correlation with the real data, it would be a leap to consider this as evidence for the existence of mechanisms operating in this manner.

It is likely that any additional signal at a different wavelength (or even with a different frequency component) would have been able to deliver a higher correlation, since the data is noisy and through the random recombination we are functionally allowing every option to be tested. This is highly likely to result in over-fitting. One way to test whether this has occurred is to plot the contributions of the top performers from these situations (analogous to Figures 4.25 and 4.26) and consider the apparent trends. This is plotted in Figure 4.28. It can be seen that there do appear to be some trends for both rods and ipRGCs. Splitting this apart, into Figures 4.29 and 4.30, where we plot the results for simulations run where only the rods were added or only the ipRGCs were added, we can see these trends slightly more clearly.

Figures 4.29 and 4.30 appear somewhat similar. Considering the similar spectral characteristics of rods and melanopsin, it should be expected that the model would use the signals in somewhat similar fashions. It appears as though neither
4.3. Results

Figure 4.28: As per Figure 4.25 but for the conditions where both rods and ipRGCs were included as additional input signals were considered.

Figure 4.29: As per Figure 4.25 but for the conditions where rods alone were included as additional input signals.
Figure 4.30: As per Figure 4.25 but for the conditions where ipRGCs alone were included as additional input signals.

has a particularly strong role to play in the L and M adaptation, however both seem to reliably be used as a negative weighting for S. It is unclear whether this is simply a response to a single high datapoint in this dataset (the peak at 460nm) or whether this integration serves a broader purpose. This peak at 460nm also seems to be the cause for the stubbornly low correlation coefficients for the S channel (compared to L and M), which max out at around 0.81 (see Figure 4.27).

4.4 Conclusion

Two methods of analysis for the MacDonald and Roque [205] data are presented here.

The chromaticity-based analysis showed that there was a strong correspondence between the patterns of the chromaticities of the adapting fields and the pattern of responses. There were a small number of outliers, which could plausibly be due to additional inputs to the adaptive process, however these outliers were in line with the amount of noise in the data, and confounded by many other variables
4.4. Conclusion and sources of noise. This analysis provides no basis for rejecting the null hypothesis that cones and rods are the sole responsible agents in adaptation for the studied retinal locations and conditions.

The spectrum-based analysis showed that the results for one observer could not be well fitted by a simple model of Von-Kries-type observer, but that simple linear combinations of the responses of such an observer could be made to fit the recorded data very well. Though the ability of these fits is to be expected from this type of post-hoc fitting, the types of models which are predicted by the fitting align well with our understanding of post-receptoral signals, which suggests that we may be measuring adaptation at these levels. It would be particularly valuable to see whether the temporal nature of responses also aligned with our understanding of the timecourses of these different signals. It should be noted that this result could be due to the nature of the experimental set-up; observers were unable to modulate the cone activations directly. For example, to make the stimulus more green, the observer would inherently have to make it less red. It is plausible that this may account for the relationships seen in this analysis.

The second analysis further showed that integration of the simulated rod and melanopic signals provides additional value in fitting the data. Again, this is to be expected from this type of fitting. The minimal improvement delivered by the inclusion of rod and melanopic signals does not furnish me with strong enough evidence to reject the null hypothesis that chromatic adaptation can be fully accounted for by cone and rod mechanisms.

4.4.1 Limitations

Several limitations have been identified with this experimental design which should be borne in mind when analysing this dataset, or planning similar experiments.

- The light levels in this experiment were in the mesopic range, and it is unclear whether we might expect to see melanopsin activation, and thus any melanopic interaction to adaptation, at these levels.

- The light levels were different for each condition. Presumably a stronger
adapting illumination might have a stronger adaptive effect, but it is unclear what how this might manifest (faster? more chromatic neutral point?) and what the underlying relationship may be. It would be tempting to match the adapting surrounds for luminance, or radiant power, but neither of these solutions provides genuine equality across conditions (when matching for luminance almost no level of 400nm radiation would be able to match the luminance at other wavelengths, and there is no reason why radiant power should directly translate to adaptive effect).

- This experimental paradigm requires the assumption that adapting one part of the retina (the periphery) has an effect upon the adaptive state of another part (the fovea). This assumption is implicit in all chromatic adaptation models, but to some extent we know this to be incorrect - consider for example spatially locked after-images. See also the previous work of MacAdam [204] where two halves of the retina were explicitly adapted differently.

- This method provides very noisy data, and it is difficult to average over any of the nominal ‘repeat’ conditions since every variable seems to have a non-negligible effect, and none of these effects appear to be easily modellable. The inherent noise in this data collection methods places heavy limits on the scale of detectable effect sizes.

- It is difficult to distinguish effects that indicate a biological basis and those which arise due to experimental limitations. For example, in the spectrum-based analysis it was unclear whether an increase in one sensor activation was required to make a match, or whether it was simply yoked to a decrease in another by the restraints placed upon the response space.

- It appears that there is a strong correlation between an observer’s match and the previous match. It is unclear whether this is due to the fact that the new ‘random’ starting condition is centred upon the previous selection, or whether this is a foveal adaptive effect.
4.4. Conclusion

- The spectral resolution of 20nm is relatively low for analyses such as the spectrum based analysis. However, increasing the resolution would probably be an unrealistic goal, considering the amount of required observer time.

4.4.2 Further Work

This dataset may only comprise data from two observers, but it is broad and may be valuable to those interested in chromatic adaptation and colour constancy. Various further analyses are envisioned:

- As previously mentioned, there would likely be some value in the further modelling of the potential response space of an observer on this task, and the interaction the types of analysis performed here.

- It would be interesting to consider non-linear adaptation responses. This may better account for the data at the extremes of the wavelength range where luminance was very low.

- The temporal dimension of the data is not considered in either of the analyses presented here, but the code for the second analysis has been written in such a way that it should be relatively easy to implement. Considering the different expected time courses for adaptation for cones/rods/ipRGCs this may serve to be a fruitful avenue.

- It would be valuable to collect data for more observers, ideally under conditions where the white point of the display was matched between observers, in order to understand what effects are robust across observers.

- If the L* dependent effects could be accounted for, then averaging over the entire range of L* could be implemented, which should reduce the level of noise, and improve the ability of an investigator to draw conclusions from a chromaticity-based analysis.

- For the spectrum-based analysis, currently only a single colorimetric observer is used (Stockman-Sharpe 10deg). It would be possible, and more correct to
use specific observers relating to actual ages and visual fields. It may also be possible to use sharpened spectral sensitivities (see Finlayson et al. [109]), though this would need to be done particularly carefully, considering the already large potential for overfitting.

4.5 Interim Summary

The experiment reported in this chapter aimed to extend our understanding of colour constancy and chromatic adaptation, specifically asking whether there was a melanopic influence. No clear effect for a melanopic influence was found, though the absence of an effect could not be authoritatively confirmed.

A large number of limitations were identified, and it was deemed appropriate to develop a second version of this experimental set-up, and perform a further experiment. This further experiment is reported in the following chapter.
Chapter 5

Small Sphere Experiment

The work presented here has been presented as an invited oral presentation by Lindsay MacDonald at AIC 2017 [206].

5.1 Summary

This experiment was performed to develop upon the Large Sphere experiment (Chapter 4) by narrowing down the number of variables and more directly exploring the question of whether melanopsin plays a role in colour constancy. Observers were adapted to one of two perceptually metameric conditions, one with a high melanopic content and one with a low melanopic content.

It was predicted that if ipRGCs played a role in chromatic adaptation or colour constancy that we would see different achromatic settings for the mel-low and mel-high conditions. No prediction was made regarding the magnitude or direction of the effect.

Statistically significant but low magnitude differences were found between conditions for two out of three observers. For the two observers who performed repeated trials, inter-trial variability was high, and of a similar magnitude to inter-condition differences. One observer provided drastically different responses during the repeat sessions; hardware issues which may cause such a difference are discounted, and remaining hypotheses for what may have caused such a large distinction are discussed.

1The proceedings are not currently available online, though a note at https://aic-color.org/page-18077 suggests that they will soon be available.
This study was approved by the UCL Ethics committee (Project ID Number: 9357/003), application attached as Appendix B. Code and data are provided: https://github.com/da5nsy/Small-Sphere.

5.2 Introduction

Several key alterations were made to address limitations of the Large Sphere experiment:

1. Instead of 16 surround conditions, only two were included.
   
   (a) These two conditions were generated through use of narrow-band LEDs rather than filtered white light, designed to be perceptual metamers for individual observers, but with maximally different levels of melanopic activation.

2. Instead of 16 lightness conditions, only 5 were included. This was done so that more repeated conditions could be measured within the same time-span.

3. The sphere used was smaller, and the inner surface was painted with a higher reflectance paint, which raised the level of illumination to roughly 10 cd/m^2.

Further details on all of the above amendments will be included in the following sections. Three observers were tested (the author and two colleagues of a similar age), with repeats of each condition. The null hypothesis that this experiment aims to test is that melanopsin activation does not alter an observers’ perceptual white point.

5.3 Materials and Methods

5.3.1 Hardware

The sphere used in this experiment was 400mm in diameter, with ports of similar functions to those in the Large Sphere. On one side there was a padded port for an observer’s face. Mirroring this was a circular aperture of 52mm diameter (giving a viewing angle of 6.6 degrees), through which an LCD screen was visible. At the
top of the sphere was a port through which adapting illumination was provided. An additional port, on the observer’s side of the base, was added such that the illumination provided to the sphere could be unobtrusively monitored throughout experiments. A schematic is shown in Figure 5.1, and a photo of an observer in position for the experiment is shown in Figure 5.2.

**Figure 5.1:** The small sphere set up, reproduced from MacDonald et al. [206], courtesy of Lindsay MacDonald.

### 5.3.2 The Sphere

The sphere used in this experiment was smaller than that in the previous experiments (hence the experiment short-hand names) and this served several purposes. The illumination in the Large Sphere had been very low, partly such that rod interactions would be made visible, and partly due to practical limitations. The grey paint on the interior of the Large Sphere was chosen to limit specular reflections, but it meant that overall illumination levels were very low. In the small sphere, it was hoped that by reducing the size of the sphere it would be possible to increase the level of illumination, as it would be spread across a smaller surface. Additionally, of practical concern, it was easier to find experimental space for a smaller sphere.

Several paints were trialled for use in the small sphere. The required conditions were that they were available as a spray, and provided as ‘matt white’. Sample
patches were sprayed upon a piece of opaque perspex to assess finish and spectral reflectance. It was seen as beneficial if: the finish had a very fine grain and as little gloss to it as possible, the surface reflectance was high, and the SRF was as uniform across the spectrum as possible. It was also considered a requisite requirement that any paint should not include any fluorescent whitening agents (such as can be seen in the ‘white paper’ shown in Figure 5.3).

Measurements of the SRFs of the 8 tested spray-paints, and further details of the spray-paints themselves, can be seen in Figure 5.3. Measurements were made with a PR650, illuminated at 45° and measured at 90°. It can be seen that Montana Gold Sh. White Cream and Pebble, along with MTN 94 RV-198 are either too low in reflectance or not uniform in spectral reflectance across the spectrum. The two with the most desirable finishes were the Flame Blue and MTN Water Based paints.
The MTN Water Based paint was chosen due to its particularly fine-grain finish.

5.3.3 The Screen

The screen was the same screen as used in the Large Sphere experiment, but was offset by roughly 150mm through a short black paper tube in order to limit the interference (either way) between the screen emission and the sphere illumination. This was seen as a great improvement over the Large Sphere experiment, where no such effort was made. Analysis showed that at the average pixel values of the selections made by observers there was negligible screen chromaticity difference between the case where the internal sphere illumination was on and where it was off. However, at levels in line with the darkest matches made by observers, it was found that there was clear, but relatively minor, intrusion. The level of this intrusion is illustrated in Figure 5.4.

5.3.3.1 Characterisation

Following issues potentially stemming from uncontrolled variations of display output during the Large Sphere experiment, characterisations of the display output were performed more regularly (after every set of observations), and an additional characterisation protocol was introduced. The adapting illuminant was also monitored throughout every session.

The primary characterisation protocol was as standard - a spectral measurement of the display primaries followed by a ramping through intensity for each display channel from zero output to maximum output. This was only measured for the small portion of the screen which would be seen during experiments.

The secondary characterisation protocol attempted to detect any issues which may arise due to the specific way in which screen output was specified and recorded during the experiment. The main MATLAB experimental script was modified to

---

2This intrusion was only found after data had already been collected and so no attempt was made to rectify this situation. In future experiments a more effective foil might be made by using a longer tube, either folded in the manner of a traditional light trap, and/or lined with a less reflective material such as velvet.

3Controlled by the script available at https://github.com/da5nsy/Small-Sphere/blob/5c6af38c5036a4c0a328a9854427ae8e851e84fd/Hardware%20Specs/PR650%20Screen%20Measurements/PR650displaycharacterisation DG.m
Figure 5.3: The SRFs of the 8 tested paints (relative to the white patch of a MacBeth Colour Checker card) with an additional measurement of a piece of white office paper to demonstrate how a fluorescent additive might appear. Data available: https://github.com/da5nsy/Small-Sphere/tree/master/Hardware%20Specs/WhiteSprayPaints
5.3. Materials and Methods

**Figure 5.4:** Chromaticity values measured from the screen during a number of characterisation procedures. All channels were set a pixel value of 25, in line with the lowest value that an observer picked as an achromatic match, as this is where the effect of intrusion of the light from the sphere to the screen was likely to be greatest. The black asterisk represents the dark condition (no internal sphere illumination) and each red point represents one of the primary characterisation measures (with internal sphere illumination as it would be during each experimental session).

provide a ‘characterisation mode’, with a greatly reduced number of trials (15 total). This mode otherwise performed in exactly the same way that the main experimental script normally would, and the observer was replaced by the PR650, and no effort was made to select neutral points, with a measurement being made of each presentation directly. In this way, a random selection of points were recorded, and through comparison with the recorded ‘responses’, it could be seen whether there was any discrepancy between the expected and actual display output.

### 5.3.4 The LED Rig

The lighting in this experiment was designed such that two lighting conditions could be defined for each observer which were perceptual metamers (in the periphery, at
high temporal frequencies), but which differed in melanopic activation.

Four types of LED, to operate as two pairs, were chosen to maximise melanopic contrast between the pairs. An additional limitation was that no LED with a peak wavelength shorter than 400nm was to be used (due to safety concerns).

50 LEDs, mounted in a breadboard, were controlled by an Arduini Uno. Their SPDs are shown in Figure 5.5.

- 20: Bivar UV5TZ-400-15 (henceforth 'UV')
- 10: Cree C503B-BCS-CV0Z0461 (henceforth 'blue')
- 10: Cree C503B-AAS-CA0C0251-015 (henceforth 'amber')
- 10: Cree C503B-RAS-CY0B0AA2 (henceforth 'red')

![Figure 5.5: The normalised SPDs of the Small Sphere LEDs.](image)

The arduino script set the LEDs, via pulse width modulation, to the output levels decided in the perceptual nulling segment of the experiment. The two modes
5.3. Materials and Methods

had either the combination of UV and amber, or blue and red, allowing for the chromaticities falling upon the lines shown in Figure 5.7. The combinations UV and blue, or amber and red, were never used.

An example of the SPDs of the illumination inside the sphere under different conditions (following perceptual nulling, described in Section 5.3.5.1) is shown in Figure 5.6, for both conditions for observer HC. These conditions were perceptually metameric but had melanopic contrast of 309% (for this observer).

![Figure 5.6: The SPDs of the two conditions for observer HC.](image)

This breadboard was mounted stably atop the sphere with a perspex diffusion sheet between it and the opening into the sphere. Black card was used to mask stray light, mainly to address the specific concern that light may fall onto the LCD display.

Throughout the experiments the illumination inside the sphere was measured with an Ocean Optics USB 2000+ through an optical fibre probe mounted to the base of the sphere and directed at a point on the roof of the sphere, opposite the
Figure 5.7: The chromaticities of the Small Sphere LEDs in CIE 1931 space, with lines connecting the pairs which were activated simultaneously. From each pair, illumination with chromaticity values extending along each line was generable. Even if there were differences in the spectral sensitivities of the observers (compared to the CIE 1931 observer), with these narrow-band primaries there should theoretically always be a point at which the two lines cross.

observer. It was expected that the output of the LEDs may change slightly over time.

The melanopic contrast between conditions, in terms of Weber contrast, was 365%, 309%, and 488% for observers DG, HC and LW respectively. The variability in these figures between observers is due to the unique settings that each observer made, described in Section 5.3.5.1.

5.3.4.1 Light safety

Since illumination of wavelengths shorter than 400nm would be present, it was deemed appropriate to take extra precautions to ensure the safety of participants. The implementation of the ‘BS EN ISO 15004-2:2007 Ophthalmic instruments - Fundamental requirements and test methods - Part 2: Light hazard protection’
[162] provided in the Silent Substitution toolbox⁴ created by Spitschan et al. [281] was used to evaluate the safety of the experimental set-up, and under the default assumptions for pupil size, and with the maximum output from the LEDs, the stimulus was found to be considerably below the limits for type 1 continuous wave instruments. See Appendix B for further details.

5.3.5 Observer Task

In this experiment there were two stages to the observer task; the first (perceptual nulling) where observers would set the two surround conditions such that they were minimally distinguishable for them, and the second which was an achromatic selection task (much like that performed in the Large Sphere experiment).

5.3.5.1 Perceptual Nulling of Peripheral Adapting Field

The basic logic of this experiment is as follows: under the null hypothesis two adapting fields of identical appearance should cause an observer to be adapted in exactly the same way. In order to design two adapting field illuminants which appear identical (perceptual metamers) we can either make predictions based upon standard observers (with parameters set to match our real observers regards age and pupil dilation etc.) or we can employ a process whereby individual observers make minor alterations to two conditions (that are designed such that they could be metamers) until they appear identical. We have opted to use colorimetry as a starting point to choose LED primaries (Figure 5.7), and then allow observers to fine-tune this matching.

We run the risk of falling into circularity here: we ask observers to set two fields such that they appear identical, and then (in a roundabout way) we ask them whether there is any visual difference between them. If melanopsin does have a direct impact upon visual perception we are at risk of accounting for this at this stage. In an attempt to avoid this problem, we perform the perceptual nulling under conditions which we predict should not allow for ipRGC involvement, or should minimise such. It seems to be the case that ipRGCs do not react strongly over very short timescales, as discussed in Section 2.2, with cones being much more active in

⁴https://github.com/spitschan/SilentSubstitutionToolbox
this temporal window, so we chose to alternate rapidly between the two conditions and ask observers to make alterations until there is a minimal visible flicker.

Observers were instructed to fixate upon a small fixation point displayed at the centre of the otherwise dark display. They placed their hands upon three dials which could be independently varied to change the peripheral adapting illumination inside the sphere.

Two modes were presented to observers; one where the two conditions were alternated at 30hz for 1 second followed by a 10 second break, and one where the conditions were alternated at 4hz for 500ms followed by a 1 second break.

In the first mode the observer was requested only to alter the setting of the first dial, which would change the overall level of the red/blue combination (whilst the uv/amber combination remained at the same level). This was referred to as brightness matching.

During the second mode the observer was requested only to alter the settings of the second and third dials. The second dial would alter the relative contributions of red vs. blue to red/blue condition. The third dial would alter the relative contributions of uv vs. amber to the uv/amber condition. These options can be thought of as traversing the straight lines in Figure 5.7. This was referred to as colour matching.

The observer was allowed to switch back and forth between these two modes until they were happy with the match. Once an observer had indicated a match, the LED drive values were recorded and these were used for the observer in future sessions.

Observers found this task rather arduous, predominantly due to the inherent difficulty in making colour matches in the periphery. An additional difficulty was that since eye movements were not strictly controlled, if an observer were to look away briefly from the fixation they would be able to see the adapting field with their foveal vision. This was problematic since in most cases the peripheral matches induced strong contrast for foveal perception. Though it was unintentional, this seems to be a particularly effective way of generating the perception of a Maxwell
Further to this, the authors’ experience was that the visible periphery could be further divided, by eccentricity, into two or possibly three areas where a match in one area would not provide a match in both areas.

Despite these difficulties, observers generally found settings which for them resulted in a metameric match (where they could no longer perceive flicker, or perceived very little flicker). It should be noted that whilst perceptual matches were found, it should not be expected that these matches also be colorimetric matches for either of the CIE 2° or 10° observers, as these matches are being made at greater eccentricities than those represented by the CIE standard observers, and it can be expected that the functional spectral sensitivities of non-foveal receptors differ substantially from foveal and parafoveal receptors, if only due to pre-retinal filtering. The perceptual differences for foveal vision were large, as evidenced by the appearance of Maxwell spots. The chromaticities for the chosen illuminations are plotted in Figure 5.12.

One observer who initially agreed to take part in this study withdrew at this stage, partly due to a difficulty completing this task, but mainly due to a claustrophobic reaction. Another potential observer withdrew at this stage, since he was unable to make a match using this set of primaries, which we tentatively attribute to incipient cataracts.

This method has since been further developed by Allen et al. [4] who used a 2AFC task instead of a method of manual adjustment to pinpoint areas of perceptual metamerism.

5.3.5.2 Achromatic Selections

3 observers took part in the main experiment (the author (DG), HC, and LW).

The task was identical to that performed in the Large Sphere experiment; using two sliders (controlling the yellow/blue component and the red/green component) to set the foveal stimulus to appear achromatic. The instruction was given to set the central disc such that it appeared not red, green, blue or yellow. Once happy with their selection the observer was to press a button, at which point they would be presented with a new random colour. The same starting condition as for the
large sphere experiment was used.

As previously noted, instead of 10 runs of $L^*$ 85 descending in 5$^*$ increments to $L^*$ 10, there were 30 runs of a pseudo-randomly permuted (each run) set of stimuli specified such that there was one at every 10 $L^*$ interval between 30 $L^*$ and 70 $L^*$. The range was reduced so as to minimise the impact of gamut-boundary issues. The $L^*$ interval was increased to allow for a greater number of repetitions of specific values within a similar time frame. The order was pseudo-randomly permuted to avoid any trend based effects.

Two observers (HC and LW) performed 4 complete runs each, 2 under each condition. The author completed 2 runs, one under each condition. Each observer only completed one run per day. There was a gap of roughly three months between the initial runs of HC and LW and the repeat runs.

5.3.5.3 Data Analysis

Following calibration, for each observer under each condition, the means and standard deviation were computed for each dimension of CIELAB.

The data were submitted to a two-dimensional two-sample Kolmogorov-Smirnov test\(^5\) to compare the two distinct conditions for each observer (including data for repeats where performed). To provide a baseline-check, the same test was applied to the repeated data for like conditions, with the assumption that this would not return a significant difference.

These results were considered alongside measurements of LEDs taken during the experiments.

5.4 Results

5.4.1 Primary Data

A summary of results, where each run is represented by a standard deviation ellipse, is shown in Figure 5.8. This data is summarised numerically in Table 5.1. Breakdowns for each observer, showing each achromatic setting are shown in Figures

\(^5\)Using the function available from https://uk.mathworks.com/matlabcentral/fileexchange/38617-kstest2d-x1-x2-alpha.
Figure 5.8: A summary of all primary data in CIELAB. Ellipses represent 1 standard deviation in the primary and orthogonal axes. The legend lists observer (‘DG’ / ‘HC’ / ‘LW’), followed by condition (‘AU’ - amber / UV, ‘RB’ - red / blue) and date in M:D format.
All combinations of sets of data were submitted to a two-dimensional two-sample Kolmogorov-Smirnov test\(^6\). There was a statistical difference for all combinations \((\alpha = 0.05)\) except LW-AU-7-20 v. LW-RB-7-21, and LW-AU-10-11 v. LW-RB-10-12. With Bonferroni correction to account for multiple tests \((\alpha = 0.05/45\) \(^6\)Using the function available from https://uk.mathworks.com/matlabcentral/fileexchange/38617-kstest_2s_2d-x1-x2-alpha.\)
Figure 5.10: As per Figure 5.9 but for the data of observer HC.
Figure 5.11: As per Figure 5.9 but for the data of observer LW. Note the interaction with the gamut boundary, most easily seen in the top right of the middle subplot, where the lines in $L^*$ break down.
<table>
<thead>
<tr>
<th></th>
<th>Mean L*</th>
<th>Mean a*</th>
<th>Mean b*</th>
<th>SD L*</th>
<th>SD a*</th>
<th>SD b*</th>
</tr>
</thead>
<tbody>
<tr>
<td>DG-AU-7-21-</td>
<td>56.95</td>
<td>0.68</td>
<td>−8.96</td>
<td>15.48</td>
<td>3.37</td>
<td>5.16</td>
</tr>
<tr>
<td>DG-RB-7-20-</td>
<td>56.84</td>
<td>5.87</td>
<td>−2.26</td>
<td>15.54</td>
<td>4.31</td>
<td>7.16</td>
</tr>
<tr>
<td>HC-AU-7-21-</td>
<td>56.98</td>
<td>0.40</td>
<td>−11.96</td>
<td>15.54</td>
<td>8.22</td>
<td>12.52</td>
</tr>
<tr>
<td>HC-AU-10-18</td>
<td>56.90</td>
<td>−2.24</td>
<td>−7.17</td>
<td>15.42</td>
<td>10.85</td>
<td>13.62</td>
</tr>
<tr>
<td>HC-RB-7-20-</td>
<td>57.39</td>
<td>6.74</td>
<td>−12.21</td>
<td>15.47</td>
<td>8.68</td>
<td>11.20</td>
</tr>
<tr>
<td>HC-RB-10-19</td>
<td>57.05</td>
<td>3.71</td>
<td>−2.55</td>
<td>15.40</td>
<td>8.84</td>
<td>10.79</td>
</tr>
<tr>
<td>LW-RB-7-21-</td>
<td>57.51</td>
<td>3.39</td>
<td>20.46</td>
<td>15.47</td>
<td>10.09</td>
<td>12.43</td>
</tr>
<tr>
<td>LW-RB-10-12</td>
<td>57.06</td>
<td>44.96</td>
<td>10.44</td>
<td>14.93</td>
<td>13.42</td>
<td>11.13</td>
</tr>
<tr>
<td>LW-AU-7-20-</td>
<td>57.54</td>
<td>3.60</td>
<td>17.57</td>
<td>15.42</td>
<td>8.35</td>
<td>10.79</td>
</tr>
<tr>
<td>LW-AU-10-11</td>
<td>56.62</td>
<td>43.21</td>
<td>13.54</td>
<td>14.85</td>
<td>14.96</td>
<td>10.69</td>
</tr>
</tbody>
</table>

Table 5.1: Summary of Small Sphere data.

= 0.0011), HC-AU-10-18 v. HC-AU-7-21- additionally fell above the α threshold.
5.4.2 Secondary Data

Data describing the characterisations of hardware follow. Figures 5.12 and 5.13 show the chromaticities of the adapting surrounds as recorded during each session. Figure 5.14 shows the recorded gamut and white points of the display measured before or after each session. Figure 5.15 shows representative results from one of the secondary characterisation sessions.

Figure 5.12: The CIE 1931 chromaticities of the adapting field as measured during data collection, with the spectral locus for context. See figure 5.13 for further detail.
5.4. Results

Figure 5.13: As per Figure 5.12 but with different scaling and additional labels.

Figure 5.14: The display gamut at multiple pixel output levels, and white points, for all sessions. Measurements are grouped by pixel value (0-255).
Figure 5.15: Results from two secondary characterisation sessions. There appears to be a minor rotation in chromaticity space. Importantly, the distortion appears systematic across sessions, and therefore is unlikely to be the cause of the large difference between the two sets of LW data. The level of differences shown here mirrors that measured in all other characterisation sessions.
5.5 Discussion

There were statistically significant differences between the achromatic settings across almost all conditions (exclusions previously noted), including nominal repeat conditions. Most of these differences were relatively minor in magnitude, and within the range of what might reasonably be expected from various inherent sources of error and/or noise.

For observers DG and HC there was a consistent shift to higher values of $a^*$ under the RB (mel-high) condition, which corresponds to an NCS value of R20B (red with a small amount of blue) according to Derefeldt and Sahlin [78]. In the case of observer DG this was accompanied by an upward shift in $b^*$, which suggests a shift to a yellower red.

For observer LW there were not significant differences between the different conditions, but there was a large magnitude difference between the repeat conditions (See Figure 5.8 or 5.11). To be clear - with a single day separating inter-condition trials, this observer exhibited no difference between conditions, but after roughly three months, when returning to do a second pair of trials (again with a single day separating) the results were strikingly different. This is particularly unusual considering how well matched the inter-condition responses are; if it were just the case that this particular observer was particularly unreliable in their selections (and their data does exhibit generally quite high standard deviations, see Table 5.1) then I posit that they would be as unlikely to be able to repeat their selections after a break of a day as they would be after a break of three months.

The differences cannot readily be explained by hardware issues; variation in the surround illumination and screen output are minimal between sessions. Further, the same level of hardware variation would have been present for both observer LW and observer HC, and no such dramatic shift is seen for HC’s data.

Regarding Figure 5.13, which shows the chromaticities of the peripheral illumination, it can be seen that there was some variation between repeated conditions, and also a small amount of variation within each session. The variation across sessions is likely due to physical disturbance of the equipment (either the measure-
ment device or the illumination source) and is of a relatively minor magnitude. The variation within session seems to be systematic, perhaps relating to warm-up time of the illumination source, with the chromaticity being quite variable for the first few minutes of several trials before gaining stability.

Figures 5.14 and 5.15 show the results of the primary and secondary characterisation protocols respectively. Figure 5.14 summarises all the collected data, showing that at very low pixel values there is a moderate level of variability, which seems to correspond to the sphere illumination, suggesting that some illumination from the sphere does reach the screen. Figure 5.14 also shows the recorded white point of the screen for each characterisation, but these points stack fairly precisely atop one another. Figure 5.15 shows two representative sets of results from the secondary characterisation protocol (for observer ‘LW’, under red-blue adapting illumination, both trials). It can be seen that there is some distortion between the values as recorded (following calibration) and the measured screen output. There are several potential sources for this error; the calibration routine (for the calibration we rely on tristimulus values computed directly by the PR650, whereas for the measurements we record a spectrum and then compute our own tristimulus values), the experimental code (there are a large number of colour-space conversions in the experimental code, it is possible that one has a minor error), amongst others. Importantly, the distortion seems to be systematic (a slight rotation of points roughly around the white point) and of regular magnitude across sessions.

This suggests that there must be some other cause for the difference between LW’s first set of data and the repeats. Since it is not the focus of this investigation I shall not dwell upon this point, but I shall suggest two possible causes which have presented themselves.

The first - although relatively clear instructions were given to observers, it is possible that observer LW chose a distinct interpretation of instructions for the second set of measurements. Specifically, it is possible that the observer switched between making a ‘hue / saturation / brightness match’ to making a ‘surface-colour’ match. It does seem to be the case that for the second pair of datasets, the chosen
white points are considerably closer to the chromaticity that a neutral surface would possess, so much so that some responses from this observer hit the display gamut (this is what causes the degradation of data visible in the top right hand corner of the middle plot of Figure 5.11). If this were the case, it would underscore the importance of clear and intentional observer instructions (See Foster [113, p.679] and Arend and Reeves [6] for discussions of instructional effects).

The second, though I am aware that this sounds rather fanciful, is that this result could represent a genuine shift in the perception of the observer between sessions. Though it seems unlikely that such a drastic change could result from such, in the absence of other options I feel it is worth considering. There is evidence to suggest that an observer’s white point may change seasonally [319], as an effect of changing chromatic distributions of one’s surroundings. Three months between sessions seems like a reasonable time-frame within which to witness such a change, particularly when the difference is between July and October (Summer to Autumn), where it is most likely that there would have been a considerable difference in the lushness of vegetation.

One argument against this line of reasoning is that observer HC did not exhibit a similar shift over the same time-frame. It is quite possible however, that one observer had a greater level of exposure to natural (and thus chromatically changing) conditions during this time.

The data of observer HC by comparison shows a high level of regularity. For each session there is a regular orientation of variance (see Figure 5.8), and there is little variation either between conditions or between repeats. What difference there is, could tentatively be considered as systematic - for both the initial and repeat conditions, the RB condition receives responses that are offset to higher $a^*$ values compared to the AU condition. Both repeated trials have higher $b^*$ means than the initial data-sets, but this change seems to be in line with magnitude and direction of change between the chromaticities of the adapting fields between conditions (see Figure 5.13, where the for the second set of recordings the chromaticities have shifted upwards and rightwards in CIE 1931 chromaticity space - roughly towards
yellow).

The data of observer DG (the author’s data) exhibits a lower level of variance (probably due to increased level of familiarity with the test) and exhibits an inter-condition difference of similar magnitude and direction to that for observer HC.

It would be desirable to understand the effect of luminance or \( L^* \) on the achromatic settings made, but as with the analysis performed for the Large Sphere data, there would be no means by which it would be possible to separate out the effect of the experimental set-up (whereby the sliders represented higher chromaticity shifts per unit movement at higher values of \( L^* \)) from any underlying effect. For this reason such an analysis has not been performed.

Colour constancy indices have not been calculated for this data for a number of reasons. Firstly, there is no meaningful white point which could be used as a ‘pre-adaptation’ point in such calculations. Secondly, the ‘ideal match’ values (which would under normal circumstances be the peripheral adapting illuminant chromaticities) are designed in such a manner that they are metameric but also have a large difference in chromaticity space (Figure 5.13) due to the difference in SSF in the fovea and the periphery. For this reason, a chromaticity based description of the surrounds is likely to be both misleading and disingenuous, given our lack of knowledge about the SSFs of peripheral vision.

It is for this reason also that I have been careful not to plot the chromaticities of the surrounds alongside the chromaticities of the central achromatic settings.

5.5.1 Limitations

One key limitation in this study was that there was no prior prediction of the direction, nor magnitude, of the difference that we expected to see as the effect of changing the melanopic activation. This stems from the fact that though several researchers have sought to examine whether there is a link between melanopsin and colour constancy, none have proposed a clear theory for how or why such involvement might exist.

With the benefit of such a framework we would be much better placed to assess the results of an experiment such as this. As it is, we see minor (but statistically
5.6 Interim Summary

As we saw in the Large Sphere experiment, it is very difficult to fully control conditions such that a valuable repeated condition can be performed; all of the dimensions over which a repeat could be performed (time, observer, L*) seem to have meaningful but complex effects!

If there was a model which predicted a specific difference with a predicted orientation and magnitude, it may be possible to use alternative analysis methods which would provide greater clarity. Further, it may be possible to design experiments to more robustly test a more specific hypothesis. Chapter 7 aims to fill this gap in our understanding.

The other key limitation, which applies equally to this experiment and the Large Sphere Experiment, is that there is an implicit assumption that foveal adaptation is affected by peripheral stimulation. Though this is implicit in all CATs (with only a single transformation applied across an image), it is not clear whether there is a physiological mechanism by which a single transformation could be applied across the entire retina. Indeed we know, from simple experience with after-images, that adaptation to a small area of the retina is certainly possible. It remains unclear what the level of adaptational cross-talk across the retina is. To minimise the impact of this issue, and to increase our understanding of this process, it is recommended that further experiments of this type employ a modified design such that both the periphery and fovea (and further divisions) can be used as both adapting field and test field. Further proposed modifications will be discussed in the final chapter of this thesis.

5.6 Interim Summary

As with Chapter 4, the goal of this experiment was to explore whether there was a melanopic influence to colour constancy or chromatic adaptation. As with Chapter 4 no clear effect was found.

However, it was noted that there were a great deal of assumptions about how a melanopic input might operate, which were implicit in the experimental design,
which were not necessarily backed up by what we know about ipRGC physiology, or ecological requirements.

It was therefore decided that further work was required to understand the ecological requirements for colour constancy, and to work out abstractly whether a melanopic input might be of value to colour constancy. This inspired the work presented in Chapter 7. Alongside this, an experimental method was explored which would allow colour constancy experiments to be performed in a wider range of environments, including more ecologically valid ones, than had previously been used (Chapter 6).
Chapter 6

The Tablet Method

The work presented here has been presented previously as an oral presentation at AIC 2016 [122, p. 125] and as a poster presentation at ECVP 2017 [237, p. 93].

6.1 Summary

The principal aim of this piece of work is to explore a novel method for colour constancy experiments, which allows for experiments to be performed in real and complex environments, in order to better understand colour constancy outside of a laboratory environment. The Spatial Achromatic Point Setting (SAPS) method uses a tablet computer to present a spatial version of an achromatic point setting task.

The key concerns are whether the proposed methodology can: a) present a stable stimulus across disparate environments, b) record differences in the state of chromatic adaptation, and c) be suitable for naive observers following minimal instruction.

On each point, the method is shown to be moderately successful, with caveats, and suggestions are provided for improvements upon the current design. Such a methodology could be used to investigate the effect of multiple cues, conflicting cues and cues which cannot easily be reproduced in a laboratory environment.

Code and data are provided: https://github.com/da5nsy/SAPS/.

---

1 Abstract available: doi:10.6084/m9.figshare.4269680.v1
2 Poster available: doi:10.6084/m9.figshare.5476493.v1
6.2 Introduction

To further our understanding of colour constancy, investigators have traditionally designed well-controlled experiments where the number of variables is greatly reduced compared to a real-world environment. This allows investigators to carefully query the impact of any individual variable, or the interplay of a small number of variables.

For example, a common stimulus-type used in such experiments are ‘Mondrian’ patterns, arrangements of flat and unmoving overlapping coloured paper rectangles, see Figure 6.1 [154]. Consider also the experiments of Kraft and Brainard [176] where objects representing potential colour constancy cues such as a tin-foil covered cone (specular highlights) were removed from a neutrally coloured box one-by-one in order to probe their relative usefulness as cues for colour constancy (see Figure 6.2).

It is clear that these set-ups are much reduced in their complexity compared to a natural scene. Simple experimental stimuli allow for clear questions to be asked, and for those questions to be answered with statistical strength. Such research is valuable, but the results cannot always be extrapolated to other types of stimuli and different environments. The use of simplified stimuli risks overlooking unknown scene attributes, and the human behaviours which may be reliant on them. Experiments often find that colour constancy in lab environments is never ‘complete’ [113, 230]; is this representative of real world behaviour or could this result be due to a lab environment failing to deliver all the cues available to an observer in a natural environment? Recent technological advances have enabled experimenters to reproduce natural scenes with increasingly accuracy and comprehensiveness [138], but the number of variables in real scenes is practically infinite, and whilst our ability to reproduce a scene increases over time as technology develops, we only reproduce what we deem important at the time. Using a real-world environment would allow for processes to occur as they do naturally, in the environment for which these processes are presumably optimised (See Kelly et al. [167] and Shepard [269]). The ability to move away from the lab environment also allows us to run
6.2. Introduction

There are at least two clear challenges to the use of natural environments for scientific study; the inability to control target variables, and the influence of uncontrollable or uncontrolled non-target variables. The first challenge may be surmountable; depending on the variable in question it might either be controlled by force, or over time natural variability may provide the required experimental range. The second may be insurmountable but, depending on the specific situation, it may be permissible to consider uncontrolled variations simply as sources of experimental noise. Further challenges arise where connections exist between target and non-target variables, or where the influence of the non-target variables
dwarf the effect of the target variable.

One further challenge: it is rare for an experiment in a real-world environment to not intrude onto that real-world scene and change it in some way. The only true solution to this problem would be to consider unannounced observation of a natural behaviour as the only acceptable scientific method, which would be incredibly restrictive and impractical. A pragmatic compromise is to design an experiment such that it modifies the environment of the observer minimally, and to consider carefully the impact that the experimental set-up may have upon observers. This is the approach which shall be taken here.

The method presented here is a variant of the ‘achromatic setting’ method.\textsuperscript{3} The achromatic setting method requires an observer, under specific conditions, to adjust the chromaticity of an item in their visual field such that this item appears

\textsuperscript{3}For an overview see Section 2.1.4.2, or the section on ‘achromatic adjustment’ in Foster [113] and ‘Matching to an internal standard’ or ‘achromatic setting’ in Smithson [274].
achromatic. Changes in selected achromatic point (in colour space) are thought to represent general adaptive shifts. For example, an observer in an environment lit by a chromatic illuminant would be expected to pick an achromatic point which is shifted towards the chromaticity of the illuminant, compared to the achromatic point which they might select in a more neutral environment. This is often practically achieved by having the ‘object’ be an area of a computer screen, and have it controllable in two or more chromatic dimensions (such as relative amounts of unique red/green and blue/yellow).

In the SAPS method, as presented here, a tablet computer is given to an observer to hold as is comfortable to them, and upon this computer an isoluminant slice through a nominally perceptually uniform colour space is presented, from which they are requested to select (by touching upon the screen with a finger) the point which they deem to be most achromatic (the specific phrase ‘grey-est or least colourful’ is employed in order to make the task suitable for non-colour-scientist observers). The term ‘spatial’ is used since the user provides information by making a spatial selection which directly corresponds to a colour choice, where in other methods abstract sliders or knobs may be used to alter the chromaticity of a static object.

To minimise the effect of the presentation of chromatic scenes upon the viewer, which may influence an observer’s state of chromatic adaptation, this process is repeated a number of times with the area of colour space which is presented varying, through random rotation about the luminance axis, and random offsetting through both dimensions of the chromatic plane.

The rest of this chapter shall describe the method in more detail, and some experiments performed to explore the potential abilities and limitations of this method.
6.3 Research Questions and Hypotheses

Hypothesis 1: The SAPS method is suitable for colour constancy experiments

To be fit for performing colour constancy experiments, this method needs to:

A. Provide a relatively environment-agnostic stimulus. This experimental method relies on the assumption that the tablet display delivers a stimulus with identical physical properties to the observer independent of environment, with no effect of ambient illumination. In practical terms, this means that the tablet must not be affected by reflection, either at the glossy surface of the tablet display, or at reflection at any other level of the display architecture, to the extent that this has a non-negligible impact upon recorded data. If it is found that this is not the case, the tablet would need to be characterised separately for each environment.

B. Collect meaningful data. One would expect this to be indicated by small intra-observer variability (not recording a change where there is presumably none), and reasonable inter-environment change (recording a change where there is presumably a change). It would be expected that these changes would be in line with previously published results. An implicit assumption is that an observer’s achromatic point will align with the chromaticity of the ambient illuminant.

C. Additional aim - Be suitable for naive observers. This would reduce the amount of time and effort required to run experiments, and allow for the collection of data from participants less likely to be biased through task expectation. It also means that the demographic group of observers is less likely to be WEIRD (‘Western, Educated, Industrialized, Rich, and Democratic’, [34, 144, 165]),

---

4In this instance by ‘naive’ I mean non-colour-scientist (a large number of colour vision experiments have been run only with colour scientists as observers and it is possible that this has biased results) and untrained (it is relatively common for this type of experiment to require extensive training upon a colour naming system such as The Munsell System of colour, by which participants are asked to report the appearance of a test object.)
6.4. Experimental set-up and methodology 197

Hypothesis 2: ipRGC activation affects perceptual white point.

Primarily as a proof of concept for this methodology, but also to assist in answering other research questions posed within this thesis, an experiment was performed whereby observers made achromatic selections under a selection of colorimetrically metameric illuminants where there was a melanopic contrast between illuminants. If ipRGCs play a role in colour constancy we would expect to see a distinction between the responses recorded under the different illuminants. If the results of Cao et al. [38] were to be reproduced, we would expect to see a distinction along $l$ pathway isolating directions and not along $s$ pathway isolating directions.

6.4   Experimental set-up and methodology

Three separate experiments were performed. These are described in Section 6.5.

6.4.1   Participants

6.4.1.1 Selection of participants

The participants for Experiment 1 (Section 6.5.1) were the author, one of the author’s academic supervisors (KC) and a technician at the laboratory (TS).

For Experiment 2 (Section 6.5.2), 58 observers were selected randomly from museum visitors. A visitor would be approached by the experimenter (the author) and asked whether ‘they would be interested in taking part in a colour vision experiment’. Those who replied positively were verbally informed of the ethics details for the study, the principal parts of which were: that no identifying information would be recorded, that the experiment carried no risks greater than those associated with normal use of a tablet computer, that observers were not paid or otherwise incentivised to take part, and that the task would take roughly ten minutes. A verbal description of the ethics details for this experiment was provided rather than the more traditional paper version to avoid providing the observer with a white reference immediately prior to the experiment. This study was approved by the UCL Ethics committee (Project ID Number: 9357/001), application attached as
Appendix C. In accordance with the ethics approval granted for this project, people under the age of 18 were only invited to participate under the supervision of a parent/carer.

For Experiment 3 (Section 6.5.3), nine participants (5 female, 4 male, ages not recorded) were recruited from the author’s friends and family (including one academic supervisor, LM), with the hope that this would assure attendance and motivation and to take advantage of short-notice availability of the experimental space. Participants were informed of the ethics details for the study in advance of attending. Ethics approval was provided following the amendment of the aforementioned ethics application (9357/001), attached as Appendix D.

6.4.1.2 Instructions to the participants

The observer was instructed to hold the tablet such as was comfortable for them to do so, upon which the first trial of the experiment was already visible (see Figure 6.3). They were instructed to ‘touch the grey-est, or least colourful, point on the screen’. Upon touching the screen, the stimulus would be replaced with a new stimulus. This new stimulus was a new subsection of the full stimulus image (See Figures 6.4 and 6.6) which had been randomly rotated and offset (further information provided in Section 6.4.2).

Figure 6.3: A participant (the author) holding the tablet with a stimulus on screen.
Most observers seemed to find this task difficult for the first stimulus (appearing confused and often verbally expressing difficulty), but within the first few stimuli seemed to develop an increased comfort and ease with the task, resulting in a decreased response time (Figure 6.39). At this point the observer was told that there would be thirty trials. No observers dropped out mid-session. No training was provided, and no runs were excluded as training runs. One stimulus was forced to have identical rotation and offset to an earlier stimulus, in order to assess intra-observer variation (see Section 6.4.4.2).

Following the thirty trials, a secondary task was presented to observers, designed to characterise their touch input. This task features a 2x2 checker board pattern upon a black background (See Figure 6.5). Observers were instructed to touch the centre of this checker board. Upon registering a touch, a new checkerboard would be presented, of the same attributes but modulated in position in the same way as the main stimulus. This stimulus is presented 10 times. This data was later used to calibrate touch input, and to ascertain the amount of measurement uncertainty derived from touch input.
6.4.2 Stimuli

6.4.2.1 Selection of stimuli colour space

The stimulus presented to observers was an isoluminant plane (in CIE L*) through CIELUV colour space. CIELUV was chosen, over CIELAB or CIECAM02 for example, since it has an associated object colour space (CIE u’v’, see Section 2.1.1.1), which allows for comparisons between the chromaticity of light sources and the chromaticity of selections to be made\(^5\).

6.4.2.2 Generating the stimuli

The stimulus was specified as: L*: 60 uniformly across field, u*: ranging linearly from -50 to 50 from one side of the field to the other, and v*: as for u*, but along the orthogonal axis, so that the stimulus was of uniform lightness and smoothly changing hue and chroma. The full stimulus image (Figure 6.6) was 2188 x 2188 pixels; this is larger than the pixel dimensions of the chosen screen (1366 x 768, Section 6.4.3) to allow for the stimulus to be rotated freely and offset by up to a third of the image in any direction before the edge of the image is encountered.

\(^5\)Some preliminary runs of the experiment, performed at The UCL Grant Museum, used a stimulus defined in CIELAB colour-space instead, before later changing to CIELUV for the reason above.
6.4. Experimental set-up and methodology

The stimulus was specified with the above attributes in a matrix within Matrix Laboratory (MathWorks) (MATLAB)\(^6\). CIELUV values were converted to XYZ tristimulus values, with reference white set as the XYZ tristimulus values of the display at maximum white (with screen protector), as measured with an Xrite i1 device (shown in Figure 6.9). Linearisation was achieved through the use of a look-up table, computed by spline interpolation of the measured outputs at 15 pixel value increments from 0 to 255 for each channel. The stimulus was then output as a 24-bit RGB tiff, which could be easily loaded and manipulated by the psychophysical stimulus presentation software PsychoPy [248].

6.4.2.3 Presenting the stimuli

A program was written in PsychoPy\(^7\) that presents the stimulus 30 times, with random rotation and random offset in the horizontal and vertical dimensions of between -1/6 and +1/6 of the respective dimension. Thus the ‘objective white point’, where \([u^*,v^*] = (0,0)\), was always within the central third (in each dimension) of the screen. This program saved details of the stimulus offset and rotation (‘dpX’, ‘dpY’ and ‘ori’), co-ordinates of observer selected point (‘x-co_raw’, ‘y-co_raw’, and also ‘direction_raw’, ‘magnitude_raw’), and time taken to make selection, measured since last selection (‘toc_raw’).

The effect of rotation and offset was such that on each stimulus presentation the observer saw a stimulus that appeared somewhat different to the previous stimulus, but still hopefully included their perceptual achromatic point. Plotting the chromaticities present across a full run of 30 trials yields Figure 6.7, where it can be seen that the rotation results in a roughly circular spread of chromaticities, and the rotation and the offsetting together result in a gradient of likelihood of presentation which is high at the centre of the full stimulus and lowest at the edges of the full stimulus. This gamut distribution is referred to further in the text as the ‘practical gamut’.

\(^6\)Code: [https://github.com/da5nsy/SAPS/blob/21940bb6ed9d037d88e1b655f4919d73431743a/experiment/stimulusGenerator002.m](https://github.com/da5nsy/SAPS/blob/21940bb6ed9d037d88e1b655f4919d73431743a/experiment/stimulusGenerator002.m)
\(^7\)Code: [https://github.com/da5nsy/SAPS/blob/21940bb6ed9d037d88e1b655f4919d73431743a/experiment/SpatialAchromaticPointSetting_0940_Grant.py](https://github.com/da5nsy/SAPS/blob/21940bb6ed9d037d88e1b655f4919d73431743a/experiment/SpatialAchromaticPointSetting_0940_Grant.py)
Figure 6.6: Full stimulus image, from which subsections were selected and presented as stimuli. Note that no effort has been made to correct this image for printing, and so it is best considered as only a rough approximation. It may be possible to see the vertical line on the left where the sRGB gamut boundary is reached; this is further described in Figure 6.14.
6.4. Experimental set-up and methodology

Figure 6.7: The ‘practical gamut’ (shown in colours ranging from dark blue to yellow), within the device gamut (dashed lines). This plot shows the relative frequency that chromaticities are actually presented, within the device gamut (more fully described in Figure 6.9). The highest number of times a particular chromaticity can be presented is 30, since there are 30 trials in a typical run. Chromaticities falling towards the objective white point of the stimulus (where \([u^*,v^*] = (0,0)\)) are presented every stimulus, whereas those further away from the centre of the stimulus are presented less frequently. On the left-hand side of the cluster it can be seen that the gamut boundary is reached.

6.4.2.4 Limitations of the stimuli

Under the current set-up the stimulus is to some extent fixed; each stimulus is a subsection of a single complete stimulus image (Figure 6.6). This means that an observer can at no point select a chromaticity that is not present in the full stimulus, and they will less often have the option to select a chromaticity which falls towards the boundary of this stimulus space than they would be able to select one at the centre of this space, due to the nature of the rotation and offset described in Section 6.4.2.3 (See Figure 6.7).

When designing the stimulus it was assumed that the stimulus would cover
a large enough area of colour-space to allow for any selection that an observer would reasonably care to make (the stimulus would include an achromatic area surrounded by areas which under no situation would be deemed achromatic). Later analysis (for example, see Figure 6.27 where an illuminant falls far outside the practical gamut) showed that this was not the case, and that it was not unusual for an illuminant that appeared entirely neutral to fall outside of the response-space. I had underestimated the power of colour constancy!

This limitation leads to a pattern in the data which I shall refer to as ‘smearing’, so called because when the desired selection point is outside of the practical gamut, or even close to the edge of the gamut, the closest possible points which are available for an observer to select will fall upon a line between the objective white point (where \([u^*,v^*] = (0,0)\)) and the true desired selection point, thus the data is smeared between the point which an observer actually wants to select, and the objective white point. This is discussed further in Section 6.7.5, where amendments to this experimental method are proposed which might limit or remove this effect.

A visual example of this can be given by considering the results of an amended experiment where the same stimulus was presented but the question changed to ‘touch the X-est point on the screen’ where X was red, green, blue or yellow. See Figure 6.8, where the responses for chromatic questions have a much larger spread than that for the achromatic question (in grey, centre) and tend to show characteristic orientations of spread - from the outer regions of the practical gamut towards the centre. This is a more extreme case than a white-appearing light source outside of the practical gamut but it is thought that the effect upon the data would be similar.

A broader limitation, which applies to all screen-based experiments, is that regardless of the gamut of the stimulus used (here ‘practical gamut’) we are operating within a larger fixed gamut, shown within Figures 6.7 an 6.9. It can be seen in Figure 6.7 that our chosen stimulus gently nudges the left-hand gamut boundary of the screen. This can be thought of as a request for colours which are less red than is possible using this specific display.
6.4. Experimental set-up and methodology

Figure 6.8: Results from an amended experimental paradigm to demonstrate ‘smearing’. The observer was asked to ‘touch the X-est point on the screen’ where X was red, green, blue or yellow. Background dots are sub-sampled from the practical gamut, to indicate ability of the observer to select specific chromaticities, as shown in Figure 6.7. A normal run of the experiment was conducted under the same conditions at the same time to provide a baseline condition. The experiment was performed in the basement of the Chadwick Building (Section 6.4.5.3) with the author as observer.

The effect of this could be reduced by choosing a screen with a particularly large gamut. A drastic alternative might be to consider a reflective screen device (such as a Kindle or other e-reader/e-ink device) but this would immediately fail Hypothesis 1A (‘Provide a relatively environment-agnostic stimulus’) and so the device would need to be characterised within each new environment, which reduces the appeal of this method considerably.

Another limitation with this type of stimulus/set-up is that if observers were to treat the screen as an emissive device they might treat the colours displayed as unrelated to the environment, and thus simply select a moving average of the colours which had been displayed on the screen, which would eventually (over...
enough trials) average to be the nominal white point (centre) of the stimulus. This is discussed further in Section 6.7.4.

6.4.2.5 Touch input characterisation

The checker-board stimuli (Figure 6.5), included to allow for fine spatial calibration for individual observers’ touch input, was generated by running the main experimental script again but with the stimulus file path replaced with a simple pattern generated within PsychoPy.

I hypothesise that any shifts in touch input would be the result of one or both of two potential influences: personal finger offset and general hardware calibration. By the former I refer to the fact that whilst a finger is considered to be a relatively discrete unit, controlled with presumed dexterity and accuracy, the specific part of the finger which touches the screen will vary from person to person, with each person exhibiting a reliable bias. By the latter I refer to the fact that it is highly possible for there to be a ‘fairground gun effect’, whereby the hardware introduces a reliable bias due to calibration misalignment, which affects every observer equally.

Performing a calibration of each observer’s data based on this individual characterisation allows for offsetting of both of these types of variation.

6.4.3 Specification of Tablet PC

Participants undertook the experiment upon a Dell Latitude 10 ST2 tablet computer (223x126mm active screen area, 1366x768 pixels) with ‘BROTECT’ Matte Screen Protector (a thin adhesive, translucent, and matte screen protector). This tablet was chosen for its ability to run a Windows environment, and the screen protector was added to reduce the severity of specular reflections. The measured gamut of this device is shown in Figure 6.9, with the gamut of sRGB for comparison.

6.4.3.1 Impact of ambient illumination

A key requirement in this methodology is that the display device remains roughly colorimetrically stable across a range of lighting environments (see Hypothesis 1A). This would only be true if the device produced the entirety of the light emanating from it in normal use. In reality, a small amount of light will be reflected from the
surrounding environment. This may be reflected either at the surface layer of the screen (specular reflections) or at a lower level of the screen architecture. Here we aim to quantify the amount of light reflected in this way, understand the impact that this has on this method, and seek to minimise this impact if possible.

It is assumed that specular reflections are clearly distinguishable to most users, and that users will automatically hold the device in such a way as to minimise their interference with the task. It is also assumed that the spatial nature of such reflections and the fact that they are not locked to the geometry of the screen (but rather move as the screen or observer moves) would further allow an observer to visually discount them and not confuse them for an output of the screen.

The key concern then is reflection at other levels of the screen architec-
ture. Using the screen calibration framework [18], this may be considered as an environment-dependent ‘offset’, that is, a figure which is added to the output of the screen at all levels. It is assumed that this level is constant in an unchanging environment, and independent of the output of the screen. It therefore follows that this will have greatest impact on the chromaticity of the display at low luminances, where the amount of reflected light is high relative to the output of the display.

To consider whether such reflections exist for our specific set-up, telespectro-radiometric measurements were taken of screen at varying pixel value levels (0 to 255, intervals of 15)\(^8\) under 3 different lighting conditions (Warm White (‘WW’) : [\(u'v'\)] = [0.248,0.524]; Cool White (‘CW’) : [\(u'v'\)] = [0.200,0.465]), Mel-High (‘MH’) : [\(u'v'\)] = [0.201,0.467], further described in Section 6.4.5.1) One condition was repeated to assess measurement uncertainty. The measurement device used was a PR650. The set-up is shown in Figures 6.10 and 6.11.

Broadly speaking, at a pixel value of 0 one would expect that the illumination reaching the position of the observer depends almost entirely upon the illumination (entirely, if we assume that the black of the screen is uniformly spectrally reflective), whereas at maximum output (pixel values of 255 in all channels) the illumination reaching the position of the observer would be greatly more dependent upon the output of the device, hopefully with negligible influence of the illumination (so long as the illumination was below a certain threshold). The point of interest is therefore assumed to be the pixel value where the shift resulting from differing illuminants jumps from being non-negligible to negligible for our purposes.

As shown in Figure 6.12 it was found that for pixel values of 60 and below there was a considerable chromaticity variation between data from each lighting condition. This aligns with our expectation that the greatest effect would be seen at lower screen output levels.

For values above 75 (inclusive), a conservative estimate for the maximum shifts in chromaticity due to variation between light sources (considering only those sources tested) would be roughly 0.004 in the \(u'\) axis, and 0.009 in the \(v'\)

\(^8\)Code: https://github.com/da5nsy/SAPS/blob/master/auxiliary_functions/tablet%20characterization/calibration.py
Figure 6.10: The measurement set-up at Pedestrian Accessibility and Movement Environment Laboratory (PAMELA). PR 650 on right, with control PC on left, and tablet in the centre.

axis. These values are derived from visual inspection of the variation in values of chromaticity for readings taken of pixel values above 75. These figures could be considered baseline figures for classifying observed differences as likely to originate from genuine changes in observer state rather than lighting/stimulus artefacts.

It is noted that the variation between the repeated measurements, denoted ‘WW’ and ‘WW2’, is larger than might be expected; at high pixel values the chromaticities recorded under ‘CW’, ‘MH’ and ‘WW2’ converge very well, but ‘WW’ seems offset by roughly 0.005 units in a roughly north-east direction in colour space. The cause of this is unclear. The cause could be the result of ‘warm up’ (either in terms of an actual temperature dependency, or in terms of a device taking time to settle into a default operating mode after turn on) of either the screen or the spectroradiometer. It could also be the case that between measurements the angle
of the screen relative to the spectroradiometer changed and that this was the cause of the distinction. For this type of display it is well known that there are generally significant effects of angle of view. No attempt was made to force participants to hold the tablet at a specific angle, assuming that observers would naturally hold the tablet perpendicular to the line of view. Any attempt to force a specific viewing angle would have limited the accessibility of this methodology.

Now I shall consider the impact of the above finding upon our specific stimulus. Figure 6.13 shows the histogram of each channel within the full stimulus. Note that only the red channel has a significant number of pixels with values below 75 (with the blue channel having some values which come close), and so following from our observation that chromaticity was most perturbed by variations in lighting where the pixel value was below 75, this is the channel most likely to be influenced by
Figure 6.12: Chromaticity co-ordinates of measurements at different pixel values, under 3 conditions (Warm White (‘WW’) : \([u',v'] = [0.248,0.524]\); Cool White (‘CW’) : \([u',v'] = [0.200,0.465]\), Mel-High (‘MH’) : \([u',v'] = [0.201,0.467]\), further described in Section 6.4.5.1). ‘WW2’ is a repeat of the ‘WW’ condition. It can be seen that the 0 values (where the screen is effectively off) diverge greatly. They do not diverge in exactly the directions of the chromaticities of the light sources, which could either be due to measurement inaccuracy, or the base colour of the tablet may not be perfectly spectrally neutral. At higher pixel values the chromaticities start to converge at around \([u',v'] = [0.193,0.47]\).

In summary, measurements taken lead us to conclude that influence of the ambient illumination (though only in specific spatial sections of the stimulus image).

Note also that whilst the blue and green channels possess no pixels at either end of the pixel value range (0/255), the red channel possesses both, suggesting that a gamut boundary is reached at both extremes for the red channel. A large number of pixels in the red channel are at 0, with a small number at 255, representing points at which the sRGB gamut is reached. From 6.14 it can be seen the zero values are located on the far left of the stimulus, in the strongly blue/green area, and that the 255 values are located in the bottom right hand corner, in the strong red/pink area.

In summary, measurements taken lead us to conclude that influence of the
ambient illumination, for the illuminations we have tested, has a minimal effect on the presentation of the stimulus as defined in this experiment (up to 0.004u’ and 0.009v’). The greatest effect will likely be upon the sections of the stimulus where the pixel value falls below 75, as happens in specific areas of the red channel image. These chromaticities will not be often presented since they fall at the edge and corner of the full stimulus.

It would be possible to create a stimulus which did not reach the gamut boundaries, and which does not include pixel values below a certain value for any channel. However, if using the same hardware, the trade-off would be that a smaller section of colour space would be presentable, and this would have two knock-on effects; the task would be harder (the stimulus would be less saturated), and the results would be more tightly bounded (the observer would not be able to select more chromatic points). For further discussion of ‘bounding’ see section 6.4.2.4.

The measurements taken here are limited in scope by the fact that the effect of ambient illumination is likely linked to the overall level of ambient illumination; it is

Figure 6.13: Histogram for the different channels in the full stimulus image.
6.4. Experimental set-up and methodology

Figure 6.14: A breakdown of the stimulus image by channel. Here it can be seen that for the red channel, we reach the edge of the range, with a black bar being visible vertically on the left and a white corner (just about) visible in the bottom right. This translates, in the RGB image, to a bar on the left which appears to only vary in blue/green and an area in the bottom right which does not vary in red-ness. Note that the reproduction here is crude and only roughly estimates how the stimulus would appear during the experiment.

It also seems noteworthy to explicitly consider that we have assumed a linearity of sorts in asserting that there is likely to be a colorimetric effect where pixel values in any one channel drop below a threshold value. In reality, our tests show that chromaticity is affected when pixel values of all three channels drop below a threshold value, and it is a conservative assumption to assert that there is a risk when a single channel drops below this threshold value. It is quite possible, for example, that where values in the red channel drop below the threshold, if the surrounding blue and green pixels are being driven at high values, that bleed may likely that in much brighter conditions the illumination would have a greater effect on the chromaticity of the stimulus. This should be considered when assessing data collected in very bright conditions.
limit the practical impact upon overall chromaticity.

### 6.4.4 Data Analysis

#### 6.4.4.1 Processing Pipeline

The data was analysed in MATLAB using the following pipeline:

1. Data loaded into MATLAB from an excel file
2. Spatial touch points calibrated using touch characterisation data
3. Spatial data converted into chromaticity data
4. Data graded for performance and exclusions applied where appropriate (see Section 6.4.4.2)
5. Data plotted either as:
   (a) Full dataset scatter
   (b) Standard deviation ellipse/line plotting\(^9\).
   (c) Dataset mean plotting (where number of participants was large)

#### 6.4.4.2 Exclusion Criteria

In this task, a ‘good’ performance is one where it seems an observer understood the given instructions well, and was able to act upon them. Such a performance should be indicated by data which suggests that an observer was able to repeatedly select their chosen chromaticity in spite of the spatial relocation of this chromaticity during the experiment. It is expected that observers will perform with differing levels of competence, due to a variety of reasons, e.g.: level of commitment/interest, visual/pointing ability, understanding of the task. It seems reasonable to exclude all data from any observer where some of that observer’s data indicates that they have performed 'less well' than a determined threshold.

A measure of performance could be computed by measuring variance in a dataset for each observer. In an ideal situation, an observer would select precisely

\(^9\)Based on code from: https://stackoverflow.com/a/3419973
the same chromaticity on each trial, and so the variation would be 0. More realistically, slight variance is expected, due to input imprecision, fuzzy boundaries of acceptability and ‘smearing’ artefacts (see Section 6.4.2.4).

Two simple methods for assessing this variability are readily available. The first involves calculating the standard deviation of data for each observer; either considering a single chromatic axis, both axes, an average of both axes, or considering a newly defined axis such as the axis of greatest or least variability. The second involves comparing data from unannounced repeat stimuli (within each observation run trials 3 and 8 were identical, and if an observer was performing the task effectively the difference in records for these two stimuli should be minimal). This second method shall be referred to as Difference Between Unannounced Repeats (DBUR).

The first method has the advantage that it considers the entirety of each dataset; whereas the second method requires that general assumptions about the entirety of an observer’s data be made from only two data points. One disadvantage of the first method however, is that this measure would give a high value in the situation that there was a moderate or higher level of ‘smearing’, since increased spread could result from a situation where an observer was very good at selecting their chosen chromaticity, but was unable to do so since that chromaticity was not always displayed. In contrast, the second method should be unaffected by this.

A further advantage of the first method is that a non-arbitrary threshold presents itself; that of the level of variability which would result from a run of the experiment where a hypothetical observer selected the exact same physical point on the screen for each stimulus. This would represent the best possible performance of an observer who was unable, or had no interest in, following a specific chromaticity, though of course it would not include any variability attributable to a touch input (though this could be artificially simulated). The second method presents no such intuitive definition for a threshold.

Both assume that an observer’s state of chromatic adaptation remains stable across a run, which is not certain.
<table>
<thead>
<tr>
<th><strong>Standard Deviation</strong></th>
<th><strong>DBUR</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Considers entire dataset</td>
<td>Requires generalisation from only two points per observer</td>
</tr>
<tr>
<td>Susceptible to ‘smearing’ artefacts</td>
<td>Not affected by gamut boundary issues</td>
</tr>
<tr>
<td>Intuitive threshold: the SD of a ‘baseline’ observer</td>
<td>No intuitive threshold</td>
</tr>
</tbody>
</table>

Table 6.1: Comparison of exclusion criteria options.

6.4.5 Environments

6.4.5.1 UCL PAMELA

UCL’s Pedestrian Accessibility and Movement Environment Laboratory (PAMELA) (used for Experiments 1 and 3) consists of a large open space, enclosed within a light-proof warehouse, and is most often used for its customisable floor space, where small sections can be independently raised or lowered to recreate the spatial configurations of public spaces such as streets or railway platforms (see Cheng et al. [41] for example, and Figure 6.25). It is of interest to this project since it also has a customisable lighting rig comprising 44 addressable fixtures (see Figure 6.15), each with 6 independent channels (see Figure 6.16), which is controlled via a PC interface, and an additional high power white channel, which is on a separate control system.

![Image of the ceiling lighting rig at PAMELA](image)

**Figure 6.15:** The fixtures mounted on the ceiling lighting rig at PAMELA. Photo credit: Keats Webb.

For experiments reported here three illumination settings were defined - ‘Cool White (‘CW’), ‘Warm White (‘WW’), and ’Mel-High (‘MH’). SPDs are shown
Figure 6.16: The SPDs of the 7 different channels available in the PAMELA lighting rig, including the conditions ‘Warm White’ (‘WW’) and ‘Cool White’ (‘CW’).

Figure 6.17: The control panel for the LED rig at PAMELA.
for each in Figure 6.18, and chromaticities in Figure 6.19. ‘CW’ and ‘WW’ were system defaults, and ‘MH’ was defined to be a colorimetric match (for the CIE 1931 observer) for ‘CW’, but preferentially using an LED band with a peak spectrally close to the peak spectral sensitivity of melanopsin (∼480nm).

![SPDs for the colorimetrically matched illumination settings ('Mel-low' ('ML') and 'Mel-high' ('MH')) at PAMELA. 'ML' is identical to 'CW' (as in Figure 6.16, with naming switched only to match convention and make analysis clearer.)](image)

It should be noted that luminance was not matched between the conditions. Therefore, ‘CW’ and ‘MH’ could be described as colorimetric matches, but not as metameric matches. Additionally, for chromatically distinct conditions (‘CW’ vs ‘WW’), there is a confound of luminance variation. The illuminances of the three conditions, measured on a horizontal plane in the centre of the experimental space, were 235lx, 689lx and 806lx for ‘WW’, ‘CW’ and ‘MH’ respectively. In the opinion of the author, it would be advisable that if further studies of this type were undertaken, that conditions be matched for luminance in order to discount this as a variable.
6.4. Experimental set-up and methodology

The chromaticities of the light sources, measured with a UPRtek MK350 (a handheld spectrometer), were: ‘WW’: u′v′(0.248,0.524) (averaged over 4 measurements); ‘CW’: u′v′(0.200,0.465) (averaged over 3 measurements); ‘MH’: u′v′(0.201,0.467) (averaged over 8 measurements).

![Spectral Locus](image)

**Figure 6.19:** Chromaticities of defined lighting conditions at PAMELA. Note that ‘CW’ and ‘MH’ overlap.

6.4.5.2 The British Museum

The British Museum (used for Experiment 2) is a large museum located close to the main UCL campus, which exhibits artefacts of artistic, cultural and historical relevance. It is one of the UK’s most visited tourist attractions, drawing over 5 million visitors yearly [272].

Three spaces within the museum were used (with permission):

- Rooms 77/78 (‘Greek and Roman Architecture’, ‘Classical Inscriptions’), lit with fluorescent lighting (CCT = 2820K SD = 31, 83 lux SD = 4, CIE Ra = 90 SD = 0).
Chapter 6. The Tablet Method

- Room 25 (‘Africa’ - specifically the east section of the room), lit with incandescent lighting (CCT = 2509K SD = 42, 149 lux SD = 83, CIE R_a = 99 SD = 1).

- The Queen Elizabeth II Great Court (referred to as ‘GC’), lit with filtered daylight [115] (CCT = 6145K SD = 73, 7913 lux SD = 1891, CIE R_a = 83 SD = 0) during daylight hours, with additional lighting at twilight and after sunset. All experiments were carried out during daylight hours.

These three galleries are lit with different lighting technologies of distinct chromaticities, though Room 77/78 and Room 25 are both strongly yellow, as shown in Figure 6.24.

6.4.5.3 UCL Chadwick Building

Additional testing was undertaken within the Chadwick Building of UCL (home of the Department of Civil, Environmental and Geomatic Engineering), in light-tight
6.4. Experimental set-up and methodology

Figure 6.21: Photo of the author explaining the experiment to an observer at the British Museum. Permission to reproduce image gained from the member of public. Photo credit: Mona Hess.

Figure 6.22: Photo of the author performing the experiment in GC. Photo credit: Lindsay MacDonald.
Figure 6.23: Gallery 77, with fluorescent tube illumination.
Photo credit: https://sites.google.com/site/jwmuseumbibletours/all-artefacts/037-temple-of-artemis-at-ephesus

basement rooms fitted with fluorescent lighting.
Figure 6.24: Chromaticities of lighting conditions at The British Museum.
6.5 Experiments

6.5.1 Experiment 1

With the goal of understanding both the intra/inter-observer variability and intra/inter-environment variability, 13 runs of the experiment were performed by the author under the 3 illumination settings defined at PAMELA (see Section 6.4.5.1), 4 under ‘WW’, 6 under ‘CW’, 3 under ‘MH’, after 2 different lengths of adaptation (5 minutes and 30 minutes). In initial analyses no effect of length of adaptation was found, and so we group the data across these 2 conditions for analysis. Two other observers (KC and TS) also performed a number of observation (3 and 4 respectively) under a range of illumination settings.

6.5.2 Experiment 2

Extending the above experiment, particularly to the inclusion of naive observers, an experiment with 58 observers was performed at The British Museum. See Figures 6.21 and 6.22. The experiment was performed over five days, during which participants made observations in one of three gallery spaces. The three galleries used were Room 77/78, Room 25 and The Queen Elizabeth II Great Court, as described in section 6.4.5.2.

6.5.3 Experiment 3

In order to provide a case study for this methodology, and to investigate the role of melanopic activation upon achromatic settings, an experiment with 9 observers was performed at PAMELA (see Section 6.4.5.1 for details), under the 2 of the 3 lighting conditions previously defined (‘CW’ and ‘MH’), with a repeat of ‘MH’. A reminder: these two lighting conditions were specified to be colorimetrically matched for the CIE 1931 observer, but with differing melanopic lux. In this experiment, the condition previously referred to as ‘CW’ shall be referred to as with ‘ML’ (for ‘mel-low’), for clarity in order to align with current literature.

The null hypothesis for this experiment is: achromatic settings are determined solely by retinal cone catches. A corollary of this is that melanopic lux plays no role. CIE 1931 chromaticity is used as a proxy for retinal cone catches.
Table 6.2: Specifications of lighting at PAMELA.

<table>
<thead>
<tr>
<th></th>
<th>‘ML’ ('CW')</th>
<th>‘MH’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photopic lux</td>
<td>689.14</td>
<td>806.25 (117.0% of ML)</td>
</tr>
<tr>
<td>Melanopic lux</td>
<td>694.53</td>
<td>1207.61 (173.9% of ML)</td>
</tr>
<tr>
<td>CIE 1931 x</td>
<td>0.312</td>
<td>0.315</td>
</tr>
<tr>
<td>CIE 1931 y</td>
<td>0.324</td>
<td>0.324</td>
</tr>
<tr>
<td>CIE 1964 (10°) x</td>
<td>0.318</td>
<td>0.317</td>
</tr>
<tr>
<td>CIE 1964 (10°) y</td>
<td>0.317</td>
<td>0.337</td>
</tr>
<tr>
<td>CIE u’</td>
<td>0.200</td>
<td>0.201</td>
</tr>
<tr>
<td>CIE v’</td>
<td>0.465</td>
<td>0.466</td>
</tr>
</tbody>
</table>

After entering the main space at PAMELA, the space being illuminated solely by the LED rig in ‘MH’ mode, and after a minimal adaptation period (5 minutes) during which introductions and instructions were given, observers individually performed the experimental task in succession.

Once each observer had performed the task once, the lighting was changed to ‘ML’ lighting condition, and following another minimal adaptation period (5 minutes) the observers again performed the task in succession. Finally, the lighting condition was returned to the initial state (‘MH’, referred to from now on as ‘MH2’ to indicate that it is a repeated measure) and following a final 5 minute minimal adaptation period participants once again performed the task. Information about the nature of the lighting conditions was not provided to participants, though the change was noticeable (presumably due to a difference in photopic luminance and colour rendering), and observers were not informed that the third condition was a repeat of the first condition.

When not performing the task, other participants were seated facing away from the participant currently performing the task, so as not to put pressure on the participant performing the task, and to ensure that observers were not influenced by the tactics or choices of other participants. Participants were instructed not to use phones or other electronic or light emitting devices for the duration of the experiment, but were encouraged to engage in discussion with other participants. The order in which participants performed the task was decided by the author, in an arbitrary but not properly random manner. This order was then maintained for
Figure 6.25: The experimental set-up at PAMELA. Left: an observer performs the experiment. Right: other observers wait for their turn to perform the task.

Figure 6.26: The experimental set-up at PAMELA. An over the shoulder shot of an observer performing the task.
6.6. Results

6.6.1 Experiment 1

Following data processing as described in Section 6.4.4.1 (with no exclusions applied), a summary of data for an individual (the author) is displayed in Figure 6.27.

The primary result here is that there does seem to be a distinction between the results from 'WW' and the other two conditions, with the results for ‘WW’ being displaced northeast from those for 'CW' and 'MH'. The means for the ‘WW’ group and either the ‘CW’ or ‘MH’ groups seem to be offset by roughly 0.01 $\Delta u'v'$. This result is to be expected based on the difference in chromaticity between the conditions; under Hypothesis 1B one would expect to see a shift in responses such that the chromaticities selected as neutral matched the chromaticity of the ambient illumination. As the chromaticity of the illuminant is of a chromaticity north-east of the practical gamut (see Figure 6.27) it is satisfying to see that the recorded achromatic points are also in this direction from both the other datapoints and also the objective white point. If there were an effect of the ambient illumination upon the net output of the screen, such as discussed in 6.4.3.1 this would move the recorded points in the opposite direction to that observed.
A secondary result concerns the intra-observer variability. It can be seen in 6.28 that the standard deviation ellipses are large, but notably smaller than the baseline data set, suggesting that the observer was actually making conscious selections (as opposed to randomly hitting the screen, either at the same spatial position every time or at a new random position every time). It can also be seen that there is a reliable trend of the orientation of the ellipse, which could point to an elongation along the caerulean line (the line between a standard blue sky and the average chromaticity of direct sunlight) as seen by other researchers [20].

\[0.18 0.26\]
\[0.46 0.535\]
\[0.18 0.26\]
\[0.46 0.535\]
\[0.18 0.26\]
\[0.46 0.535\]

**Figure 6.27:** Results for observer DG for Experiment 1. Background dots are sub-sampled from the practical gamut, to indicate ability of the observer to select specific chromaticities, as shown in Figure 6.7. Translucent red, green and blue circles represent mean chromaticity selections per run. Red, green and blue asterisks represent illuminant chromaticities for different conditions.

It can be seen in Figure 6.27 that the chromaticity of the 'WW' illuminant falls far outside the practical gamut; there would have been no opportunity throughout the entire set of runs where the observer would have been able to select a pixel of a chromaticity matching that illuminant. It might therefore be expected that
we would see an extended smearing pattern in selections away from the centre of the practical gamut towards the chromaticity of this illuminant, as discussed in Section 6.4.2.4. This effect is indeed visible - see the extended nature of the blue ellipses in Figure 6.28. Ellipses for data under all conditions seem to have a standard orientation, but the blue ellipses seem to have a greater spread, both in this direction and overall. In this particular instance, it is difficult to separate the spread resulting from ‘smearing’ and the baseline spread along the caerulean line, since ‘smearing’ towards this particular chromaticity is lighting would align with that due to extension along the caerulean line.

**Figure 6.28:** As per Figure 6.27, but re-scaled and showing standard deviation ellipses to indicate spread of data. ‘BL’ is a baseline dataset, representing an observer hitting the spatial centre of the screen for every stimulus. It can be seen that in comparison to this baseline, there is a trend whereby selections are elongated along the positive diagonal. The results for ‘WW’ seem particularly elongated along this axis, which I assume to be the result of the chromaticity of ‘WW’ laying outside of the practical gamut. As before, background dots are sub-sampled from the practical gamut, to indicate ability of the observer to select specific chromaticities, as shown in Figure 6.7.
6.6.2 Experiment 2

6.6.2.1 Exclusions

Considering that this experiment was performed with observers selected from the general public it seems reasonable to assume that we may need to use exclusion criteria. Now that we have real data we can consider the two options presented in Section 6.4.4.2.

For Figure 6.29 'Mean SD' is calculated for each observer’s data by taking the mean of the standard deviation in the u’ and v’ chromatic dimensions. Difference Between Unannounced Repeats (DBUR) is calculated as the Euclidean distance between the two chromaticity coordinates selected for the two identical stimuli.

![Figure 6.29: The two measures of variability set against each other, for the data collected in Experiment 2.](image)

It can be seen that on both measures there are a small number of clear outliers; three points with high mean SD but low DBUR, and two points with both high mean SD and high DBUR. Using only one of these two measures would fail to pick
up some of these points, though as a single measure mean SD seems as though it
would be more effective in recognising these points.

A vertical line (the ‘logical threshold’) is plotted from the point representing
baseline data. This data is calculated by computing the results for a hypothetical
observer who pressed the precise centre of the screen for each stimulus, and the
precise centre of the checker-board on the touch characterisation phase.

As can be seen from this line, a threshold based on this measure would exclude
a large number of datasets. This is partly due to the fact that most real observers
exhibit orientation-dependent variability, with the greatest axis of variability gener-
ally being in line with the caerulean line, whereas the hypothetical data is nominally
rotationally symmetric (see Figure 6.28). It may be more appropriate then to com-
pare the SD of the hypothetical dataset to the SD of the axis of least variability for
real data. However, this could be seen as being overly lenient towards the real data.
As a compromise, and for simplicity, let’s consider the minimum value between
u’ SD and v’ SD (rather than the mean) as the representative for real data, which
shifts relative positions of the data and the threshold such that more datasets pass
this test.

This test still seems rather severe (Figure 6.30), and this is likely due to sev-
eral issues. Firstly, the theoretical data does not have any element of variability
introduced from touch uncertainty. Secondly, since the hypothetical data is firmly
centred on the spatial centre of the screen it avoids any element of aforementioned
‘smearing’ (where an observer might ideally select a point outside of the display
gamut). Thirdly, in situations where there is a high variation across both the u’ and
v’ axis, but a minimum is another dimension (in other words: strongly elliptical data
orientated roughly diagonally in chromaticity space), which seems common for this
type of data, the minimum between u’ and v’ is still going to be an overestimation
of a truly representative SD.

Whilst the second measure suggests no clear rationale for a threshold bound-
ary, the relationship between this measure and an observer’s performance follows
a strong and clear logic.
In the absence of a clear logical framework from which to draw a specific cut-off value for DBUR, I shall propose 0.04 which aligns with an apparent divide in this specific dataset, since there is a possibility that this divide represents a functional divide between observer behaviours. It also seems reasonable to exclude data with a Min SD >0.01, again with the rationale that for this specific dataset points above this boundary appear to be outliers from the rest of the group. These two cut-offs, which exclude 6 observers (\(\frac{58}{68}\)) from this data, are visualised in Figure 6.31.

### 6.6.2.2 Data

Data without exclusions is presented in Figure 6.32. Following exclusions, data is presented in Figure 6.33. The exclusions do not seem to target a specific type of data, nor do they seem to change any apparent trends or results for this specific dataset. Both before and after, we see a distinction between the ‘GC’ data and the
data from the other two locations, which may have been expected from the distinct chromaticity of the lighting. The responses for the ‘GC’ environment are fairly widely spread, but reliably falling to the left of the responses for the two other environments.

We see that the chromaticities of the ambient lighting in the first and second environments (‘77/78’, and ‘25’) are far outside the practical gamut. If we might have expected full chromatic adaptation in observers in these environments, then the results we see would represent the case where there is extreme smearing (similar to the ‘most yellow’ dataset from Figure 6.8). Within the data for these two environments it appears as though there might be two clusters, unrelated to location (an upper and a lower cluster). I hypothesise that since the ambient illumination in both conditions has a chromaticity far outside the practical gamut, that observers in both of these environments might be particularly disposed to take an approach to this task which was not originally envisaged, and which would not
be in line with the experimental goals; observers may be more likely to treat the tablet as an unrelated light source and thus select the average colour shown on the screen. Building an average over time would allow the observer to track the random offsets and rotations of the stimulus in a manner which would result in a low SD and DBUR score, but which would not represent the observer’s state of adaptation, and would instead only represent an observer’s judgement about the tablet itself.

It is worth noting that luminance in this environment was generally very high, and so there is an increased risk of the stimulus deviating from the desired colorimetry (as noted at the end of Section 6.4.3.1).

Figure 6.32: The mean settings of all observers from Experiment 2, with illuminant chromaticities for the 3 different spaces. Part of the spectral locus is visible in the top right corner.

In summary, for the data collected at the British Museum, we see a distinction between the ‘GC’ data and the data collected in the two other environments and minimal distinction between those two other datasets. This tallies well with witness-
6.6. Results

Figure 6.33: As 6.32 but with exclusions applied. Note the reduction in black points (Gallery 25) around \([0.19,0.47]\), and the reduction in blue points (GC).

...ing a distinction where lighting chromaticity is distinct, and minimal distinction where lighting chromaticities are similar. However, this method did not allow for the selection of chromaticities which would have represented a true non-spectrally-selective surface under two of these lighting conditions, and so the data for those conditions is unlikely to be a simple representation of an observer’s chromatic adaptation state. See Section 6.7.4 for a discussion of how we might interpret these results considering this complex representation.

The potential confounds in this experiment are, at minimum: observer (and associated variables), day and time, gallery space (e.g. architecture, objects on display, surrounding rooms), luminance, and lighting technology.

A subset of this data will be used in the Discussion section to comment upon the applicability of this method to naive observers (Section 6.7.3). The apparent division of the data from rooms 77/78 and 25 into two clusters (one near the objective white point, and one higher and to the right of this) will be discussed in Section...
6.7.4.

6.6.3 Experiment 3

Initial exclusions were based on the criteria decided for Experiment 2 (Min SD > 0.01, DBUR > 0.04). This is visualised in Figure 6.34. One observer, with all points above the min SD threshold was immediately excluded. The decision was made to exclude two further participants, one where one of their runs exceeds the min SD threshold, and another who is below both thresholds but still outside of the main grouping for all three runs.

![Figure 6.34](image-url)

Figure 6.34: As per Figure 6.31 but for Experiment 3 data. Triangles connecting points denote specific observers.

Following exclusions, averages for the remaining dataset can be plotted, coloured for lighting condition, as shown in Figure 6.35. There is no clear distinction between datasets collected under different lighting conditions.

Plotting data for individual participants separately, it can be seen that each participant exhibits moderate self-correlation; in successive trials, despite changes
in illumination, individuals provide data which appears to be similar across conditions, sometimes with a reliable bias per observer. See Figures 6.36 and 6.37 for examples of data from 2 participants, noting that the black ellipse (the baseline data) is the same for both observers and can be used as a visual anchor for comparison. Between these two observers it can be seen that the first reliably chooses points with a bias towards the lower left compared to the baseline data, whereas the second has a bias towards points on the right of the baseline data.

From these two figures it can be seen that, at least in this case, there is moderately large inter-observer variation. It is possible that when using a large number of observers an additional processing step may be required to offset for observer bias. In this case inter-observer variation does not seem to be masking a distinction between the conditions.
Figure 6.36: Standard deviation ellipses for observer PK.

Figure 6.37: Standard deviation ellipses for observer LM.
6.7 Discussion

In the following section I shall discuss the proposed methodology in terms of its abilities and limitations, referring in turn to the hypotheses stated at the start of this chapter (Section 6.3), and areas that I consider worthy of further consideration or development. I shall also comment on the results of Experiment 3.

6.7.1 Environment-agnostic stimulus

As discussed in Section 6.4.3.1, this specific stimulus, under the lighting conditions defined at PAMELA is only minimally affected by changes in illuminant (0.004 u’ and 0.0009 v’ for the specific illuminants tested), with greater effects only occuring at the outer edges of the stimulus space.

It is also worth noting that the effect upon collected data of an illuminant induced shift in display colorimetry would actually be opposite to the effect that one would expect to see from a change in observer adaptation state. For example, under a hypothetical blue light, which affected the stimulus by making it more blue, an observer would be inclined to select something more towards the yellow side of objective white.

It should be noted that this result is expected to break down under illumination at higher levels than that tested at PAMELA, following the assumption that brighter illumination would have an increasingly strong effect upon the screen appearance.

6.7.2 Meaningful data

6.7.2.1 Intra-observer Variability: Touch input

One insight into variability within this method can be taken by looking at the calibration data that each participant provides after the main part of each trial (Described in Section 6.4.2.5). Here they are asked to touch the centre of a checkerboard, and this data is primarily used to offset any bias introduced by the difference between where an observer thinks they are touching, and where the tablet records having been touched. A secondary use of this data, albeit with caveats, is to estimate the amount of variation introduced simply by touch imprecision; here the participant is given as-close-to an objective task as might be thought possible, and
thus any variability in the results must stem not from perceptual indecision, but rather the process of pointing and touching. There is a clear caveat; it seems likely that an estimate of touch imprecision gleaned from this dataset will underestimate the amount of ‘real’ touch imprecision, since observers have reliably (in my experience) modified their touching behaviour in this second part of the task, in the manner which might be expected of someone who is given a broad target to hit, and then immediately after is given a much smaller target to hit (they lean in, hold their finger more rigidly, and move more slowly).

To perform this analysis I shall use the data from Experiment 2, to get the largest possible sample and so as not to limit the analysis to those with a strong vested interest in this research. In Figure 6.38 I show the standard deviations based on just the calibration (checker-board) data. Note that here the units are pixels, as opposed to a chromaticity space. 4 outliers are excluded. These outliers appear to occur due to an observer accidentally double-touching the screen during the calibration phrase, and thus having one data point which is far outside the normal group. This does not affect (to a large extent) the actual calibration process, because there a median is taken, but it does have a rather strong effect when calculating the standard deviation of a set.

To consider the above information in practice, we need to perform a conversion into chromaticity space. For a simple analogy, let’s consider a pixel shift of 6 pixels (the rough average standard deviation in each dimension, based on a visual analysis of Figure 6.38). Considering that the full stimulus is 2188 pixels wide, and represents $u^*$ -50:50 (100 $u^*$ units it total), a single pixel shift represents a shift of 0.0457 $u^*$ units. Converting this to $u'$ ($u' = u^*/(13.L^*)$), gives us a value of 5.8590e-05 $u'$ per pixel, and 3.5154e-04 $u'$ for 6 pixels which is an order of magnitude smaller than the average standard deviations for the main datasets. I therefore conclude that touch imprecision will only have a minor influence on the data collected with this method, especially when an effort is made to calibrate out this bias. Further investigation may conclude that this element of measurement imprecision is not worth the effort required to correct for. It should be noted however, that by performing this analysis
6.7. Discussion

Figure 6.38: Standard deviations for the calibration datasets from the Experiment 2 data. Added for additional interest is the line of unity, which makes clearer the imbalance between SD in the x axis and the y axis - on this task an observer seems more likely to be have worse vertical discrimination. It is unclear whether this is a task specific result or a general phenomenon.

based on variance we have implicitly ignored any systemic hardware bias (say if every measurement was off by 20 pixels in the same direction, this would be corrected for in the data processing, but would not be picked up by this specific analysis.)

6.7.2.2 Intra-observer Variability: Test-retest

Our assumption is that observers would have a reasonably stable point of adaptation during their undertaking of the experiment, and that variability in the data would primarily be due a combination of fuzzy boundary of acceptability and input imprecision.

This variability is of particular interest because it impacts the power with which we are able to distinguish two distinct states of chromatic adaptation.

The data available with the largest number of repeats by a single observer
is that of Experiment 1, as shown in Figure 6.28. Here it can be seen that the
standard deviation ellipses are considerably smaller than the baseline data, and
that repeated measures under the same condition are reliably placed. The standard
deviation of the means of runs under ‘CW’ (where there were 6 repeats) is \([u',v'] = [0.0012,0.0019]\). Compared to the offset between the means between the two sets
which we think to be different (‘CW’ vs ‘WW’), offset: \(\Delta[u',\Delta v'] = [0.0061,0.0091]\),
there is shown to be roughly a factor of 5 \(1./([0.0012, 0.00190]/[0.0061, 0.0091]) =
[5.1, 4.8]\), suggesting that if our estimate for standard deviation is correct, and if
our effect size is correct and generalizable to other situations, this method should
be able to reliably detect meaningful differences in response.

6.7.2.3 Inter-environment Differences
Considering the standard theories of chromatic adaptation, and previous experi-
mental results, we would assume a difference in the achromatic settings of observers
in lighting conditions of different chromaticities. We would also assume the chro-
maticity of selected achromatic points would follow the chromaticity of ambient
illumination. In both Experiments 1 and 2 we see a distinction between environ-
ments where we might expect one to be seen. In Experiment 1 we see chromatic
selections in the direction of the illuminant chromaticity. In Experiment 2 this
also seems to occur, though the interpretation is less clear, with a large range of
responses for the ‘GC’ environment, and the ambient illumination chromaticity
being outside of the practical gamut for both the ‘77/78’ and ‘25’ environments.

In Experiment 1 (see Figure 6.27/6.28) we see a distinction between ‘WW’ and
either ‘CW’/’MH’, but no distinction between the colorimetric metamers ‘CW’ and
‘MH’ (with this lack of distinction being repeated in Experiment 3). This distinction
would likely have been larger if the practical gamut allowed for chromaticities
closer to that of ‘WW’ to be selected.

In Experiment 2 we see a clear distinction between results for trials performed
in the Great Court and those performed in either of the other environments. All
sets (except perhaps the group towards the centre of the practical gamut) exhibit
some drag towards the chromaticity of the lighting in the environment.
6.7.3 Naive observers

Naive observers (non-colour-scientists, who have only had a very short briefing and have undergone no training) generally seemed confident and able to perform the task once they had completed a couple of trials, but a moderate number of observers responded with hesitation when presented with the first stimulus. The response ‘but there isn’t anything grey here’ was not uncommon, and observers often ummed and ahhed for 20 seconds or so before committing to that first selection. The following selections seemed to flow a great deal more freely, with participants generally taking on average 3 or 4 seconds per stimulus (Figure 6.39), and about 2 minutes total (Figure 6.40).

Figure 6.39: Median time taken per stimulus across stimuli. Data is for all observers from Experiment 2. The point for stimulus 1 is not shown since this datapoint does not truly represent the amount of time taken by an observer. This is due to the fact that I often started the script running before approaching a participant, in order that I could start the observations without an observer seeing a white code screen on the tablet, and so in the case that the first potential participant I approached declined, this time could run into several minutes before a participant was even holding the device.
From a plot of standard deviations (Figure 6.41), it can be seen that whilst the participants with a large amount of experience performed well, a small number of naive participants actually performed better (assuming that SD is a valid measure of performance), whilst the remainder of participants performed almost as well with limited but notable exceptions. It is worth noting that whilst I have chosen this environment to examine (‘Gallery 25: Africa Gallery’) since it has the highest cross-over of individuals who can be classed as ‘experienced’, this choice might not be ideal since the Africa Gallery represents a situation where a large amount of ‘smearing’ may occur, since the chromaticity of the light source is far outside that selectable from the practical gamut (Figure 6.33) and this would result in artificially high values for SD.

### 6.7.4 Interacting with a large visual angle

At the start of this chapter I noted the undesirable potential for the experiment to intrude upon the observer’s visual world. When performing this experiment,
the tablet occupies a large portion of an observer’s central field of view. It should therefore be assumed that any low-level short term visual adaptation will incorporate an element of adaptation to the stimuli rather than being a pure measure of an observer’s adaptational state in a particular environment.

Regarding high level estimates of the white point, observers are aware that they are looking at a screen which they implicitly know to be an emissive device, with a white point not primarily determined by the ambient illumination. The ambiguity of the colour of the stimuli goes some way to masking the white point of the screen. If there was a clear white reference displayed on the screen I doubt we would see any distinction between environments. This is why the spatial calibration routine is run after the main trial and not before. It is possible however, that over multiple presentations of the stimuli an observer may begin to build up a representation of white which is based upon the white point of the screen. This seems particularly

Figure 6.41: Standard deviation in chromaticity for a subset of observers from Experiment 2 (Those from Gallery 25), highlighting data from the author and one academic supervisor (LM).
likely to occur when the illuminant chromaticity is far outside the practical gamut, as in Experiment 2 under the either of the strongly yellow illuminants. Under such a condition, an observer may be more likely to interpret the investigator’s question as ‘touch the grey-est, or least colourful, point on the screen within the context of what appears on the screen over time’ rather than relating to the screen in the context of the surrounding environment.

It is possible that this effect is visible in some of our data. Looking at the results from Experiment 2 in Figure 6.33 it seems as though the data for 77/78 and 25 may be split into two groups (a lower and a higher group). Looking at where these fall with respect to the practical gamut, with the lower one falling centrally and the upper one falling on the slope of the practical gamut, it is possible that this distinction represents two distinct observer approaches, with the lower group tracking the white point of the screen and the upper group attempting to select the white point of the illuminant.

To mitigate this issue, some potential solutions present themselves, though most introduce new problems or exacerbate existing issues.

The first suggestion: reduce the luminance of the display, thus directly reducing the amount of light added to the scene by the stimulus. The unwanted effect of this would be to make the stimulus more susceptible to colorimetric shift introduced by reflection of the ambient light, since this is dependent on the relative luminances of the surround and the display. Other experimenters, for example recent work reported from Anya Hurlbert’s lab (VSS 2019, ICVS 2019) used a small screen device which was physically masked from the effect of the ambient illumination. Physical masking might be difficult in a situation where touch responses are desired. Reducing the luminance of the display may have the additional benefit of making the display appear less like an emissive surface.

The second potential solution suggested is to reduce the size of the stimulus, either by reducing the area of the screen used, or by using a physically smaller display device. The unwanted effect of this, if done crudely, would be to reduce the area of colour-space from which observers could make selections, exacerbating
the ‘smearing’ issue previously discussed. If the stimulus was redefined to consider this, such that a larger area of colour-space was presented to observers (in a smaller physical space), this would reduce the precision of the data collection, since one source of imprecision (touch input) is presumed to be relatively invariant with the size of the stimulus (thus a smaller screen would make it more of a problem).

An alternative approach might be to modify the task in some way such that the observer answered as if the tablet was a reflective device. This might be achieved by physical modifications, such that the tablet-ness of the device was obscured, or possibly through a careful modification of the question posed to the observer, à la Arend and Reeves [6], such that the participant was explicitly asked to pretend that they were viewing a reflective surface.

6.7.5 Limitations of screen based experiments

Trichromatic display devices are unable to reproduce the entire gamut of visible colour. In the case of the specific display device used here, the gamut of producible chromaticities is visually described in Figure 6.9. The gamut of chromaticities actually displayed to observers, under this version of this method, is shown in Figure 6.7, and has been referred to using the term ‘practical gamut’.

As discussed in Section 6.4.2.4, chromaticities that lie outside this gamut are consequently not available as options for a participant to select, and in the current set-up which employs random rotation and offset, the colours towards the edge of the practical gamut are displayed less frequently than those at the centre of the practical gamut, which has the practical consequence that where observers may prefer to make an achromatic selection outside this gamut, or close to the edge of it, data may exhibit ‘smearing’. It would be possible to see, from the analysis of data, whether an observer was constantly selecting a chromaticity on the physical edge of the display, which might be a marker of such activity. This analysis has not been completed at this time.

It is therefore not entirely correct to say that an observer’s selected chromaticities will dutifully represent the observer’s achromatic point. Rather, it can be said that their achromatic point is likely to fall in the direction of the vector between
some objective neutral and the observed mean. There are several issues which result from this methodological limitation.

Firstly, the definition of an objective neutral point, from which such a vector could be anchored, is not a simple task. The most sensible option might be the chromaticity of the geometric centre of the stimulus (which I have referred to previously as ‘objective white’), which on average will be presented in the centre of the screen and most frequently. This choice is particularly tempting since this chromaticity, following our definition of the stimulus, is also the native white point of the display. However, as neat this may sound, this decision is still arbitrary to an extent however, as no true objective white point can reasonably be posited, and the white point of the display in this case bears no relation to anything other than a manufacturer decision.

Secondly, this issue makes the results, and therefore the method, difficult to compare to previous methods. Previous comparisons between different experiments and experimental methods in this area have been achieved through comparison of ‘constancy indices’ which describe in some way the geometrical relationship between an initial achromatic point, an adapting stimulus, and the posterior achromatic point\(^\text{10}\).

One modification to the methodology which may improve the situation, would be to render the stimuli on-the-fly (rather than calling the same stimuli for each presentation) with the observer’s previous selection having some influence on the new stimulus. A simple implementation of this would be to generate stimuli following the rule that the white point of the stimuli was defined by the achromatic point of the participant’s previous achromatic selection. In this way, the white point of each stimulus would, over trials, move closer to the participant’s true preferred achromatic point, even if it had not been present in the initial stimuli. This would, of course, only be successful if the participant’s preferred achromatic point was within the hardware gamut.

Such schemes have been used previously, and they have been referred to as

\(^{10}\text{See Foster [113] for both a description of the various constancy indices in use, and a comparison across a number of studies of recorded values of constancy index.}\)
6.7. Discussion

operating with an ‘adaptive starting rule’. Delahunt [75] describes the process:

“…I used what Brainard refers to as the ‘adaptive starting rule’. The \( a^* \) and \( b^* \) initial settings were randomized within the coordinate rectangle \([-25,25] \times [-25,25]\) centered on a reference chromaticity which was calculated as follows. For the first setting, the reference chromaticity was the white point defined by a reference illuminant. […] For each subsequent setting, the reference chromaticity was the achromatic setting made in the previous setting.”

Had this method been known of at the beginning of the project, it would have been implemented. I would recommend that if this method were developed further, integration of an adaptive starting rule should be considered.

Also of note is the methodology developed by Smithson and Zaidi [275], whereby colour boundaries are determined from categorical judgements (e.g. ‘Is this colour red or green?’). This type of method may be beneficial in detecting a change in chromatic adaptation even when the neutral point is out of gamut. It may also help to address the noisy nature of data collected through achromatic setting methods.

One further point of consideration which falls to be discussed within this section; the attentive reader may have noticed in some figures (e.g. Figure 6.8) that the selected achromatic points fall outside of the practical gamut. This of course represents a paradox, as participants should not have been able to select points outside of the practical gamut. The reason that this appears to occur is disappointingly mundane; whereas the practical gamut is calculated from screenshots of a set of real stimuli, and thus reports on what is actually delivered to the screen, the participant data is converted from spatial data to chromaticity data with no accounting for this, simply using the definition of the ideal stimulus prescribed at the very first stage of stimulus generation (before the practicalities of colour-space gamut restriction have had any effect). This correction was not originally thought necessary since it seemed unlikely that observers would pick points outside of the display gamut, but it could be added with relative ease.
6.7.6 Impact of ipRGCs on chromatic adaptation

Considering the lack of discernibility between data collected under the two conditions in Experiment 3, I conclude here that we are unable to reject the null hypothesis. There are multiple reasons worthy of consideration which could make the above finding a Type II error (false negative):

Our experimental power may be too low; considering that we have no estimate for effect size, experimental power is undetermined. However, if melanopic lux was a considerable contributor to the process of chromatic adaptation, even if I hadn’t seen a distinction in the group data I may have expected to see a distinction in the individual observer data (Figures 6.36 and 6.37).

It is possible, considering we only ran this experiment at our chosen levels of photopic and melanopic luminances, that melanopsin may only play an active role at other luminances. If further experiments of this type were undertaken, it would be wise to match for cone catches instead of relying on chromaticity as a proxy. This would also ensure matching for luminance. Considering current knowledge regarding chromatic adaptation differing lux levels shouldn’t be a concern, however it would be normal experimental procedure to equate luminance across conditions.

It is also possible that our melanopic contrast was too low - it is not an ideal comparison, but Spitschan et al. [285] found that melanopic weber contrast above 100% was needed to elicit a response (ours was 74%). Equally, it is possible that both sources contained a suprathreshold level of melanopic activation, and were equally saturating the any melanopic input to colour appearance. Before further studies of this type are performed, a clear framework for the involvement of melanopsin in colour constancy is required.

Future experimenters should also consider the introduction of a third condition which was colorimetrically different, but matched for melanopic lux, in order to improve the ability to predict expected effect size. One method to decide the amount of colorimetric difference, might be to follow the ‘splat ter’ logic of Spitschan et al. [283], whereby the chromatic difference is calculated to correspond with the maximum chromatic difference resultant between two nominally metameric conditions.
6.8. Conclusions

introduced by differences between a real observer and a standard observer.

Other potential improvements to the above experimental methodology include:

1. Repeating all conditions, as opposed to only one. This would improve an experimenter’s ability to assess repeatability.

2. Randomising the initial order of participants. Also, the question of whether to keep order (and thus adaptation time) the same for repeated trials is worthy of further consideration. Careful planning could allow for the parallel investigation of the effect of adaptation time.

3. Testing for colour-anomalous vision. On balance, for clarity, it would probably be prudent for future investigators to implement an additional colour-anomalous vision test as a pre-screening, rather than relying on this experimental method as an implicit colour-vision test.

6.7.7 Coding language

One further recommendation, for a future investigator wishing to build upon this method, or for myself should I return to it, would be to address the current situation of a split across MATLAB and Python. The current situation arose because, having started building the software in Python (specifically PsychoPy), due to the open-source nature of that project (and the implication that those outside of academia could more easily adopt and adapt the method, as well as the fact that this program would have to run on a small tablet, which I was unsure would run MATLAB) I found that my skills within Python were critically lacking when it came to generating colorimetrically defined imagery, and that my skills and the skills of those around me were much further advanced in MATLAB. I mention this primarily so that any future user might not attach any false significance to the split across languages.

6.8 Conclusions

Here I have presented a development upon the established method of achromatic point setting, which would allow colour constancy experiments to be performed in real and/or complex environments. This would allow more complex questions to
be asked, such as what cues observers use when there are conflicting cues, or cues that vary over space and time.

The method also has the advantage that it is quick and easy to explain to naive observers, meaning that a greater number, and broader demographic, of participants can be used compared to a traditional study. Naive observers do not seem to perform significantly worse than trained/colour-scientist observers.

The key tests for this methodology were that it presented a stable stimulus which was relatively unaffected by the ambient environment, and that it was able to record differences between an observer’s achromatic settings in situations where we would expect to see differences. Having explored both of these tests I conclude that this methodology broadly satisfies both (with some caveats); the colorimetry of the display was not affected in a way that would corrupt the stimuli under the conditions considered, and under conditions where we would have expected to see a distinction in response we have indeed seen one (and vice versa).

An experiment, considering whether melanopsin activation has an influence upon white point selections, found no evidence for such. Consideration was given to the factors that could have led to a type II error in this case.

The general limitations and recommendations for further development of the method were discussed, with a particular focus on the limitations due to the static stimulus and the bounding issues that this created.
Chapter 7

Computational Study

“You don’t really understand what you’ve got until you do a comprehensive model of it.”

Christopher Tyler
(Q&A session at VSS 2019)

The work presented here has been presented previously as a poster presentation at VSS 2018 [124], a poster presentation at the Visual Neuroscience Summer School (Rauischholzhausen 2018), and as an oral presentation at ICVS 2019 ².

7.1 Summary

A computational study was performed to explore whether a melanopsin signal would be useful for colour constancy in a real world environment, and to reduce the search space for future psychophysical experiments. This research was exploratory in nature (as opposed to confirmatory) with the goal being a furthering of our understanding of the problem and generation of hypotheses, rather than the confirmation or refutation of specific hypotheses [286].

It was found that an additional receptor with a spectral sensitivity different from that of the cone receptors was able to deliver a signal which could effectively be used to transform input signals to an illuminant-independent space.

The spectral sensitivity of melanopsin was found to be optimal for this task,
though only to the extent that whilst a number of different spectral positionings would provide a valuable signal, a sensor with the spectral sensitivity of melanopsin performs slightly better than the rest.

Notably, this type of algorithm makes no assumptions about scene-level attributes (in the way that ‘Grey World’ and ‘Bright-is-White’ do).

Code is provided: https://github.com/da5nsy/Melanopsin_Computational

7.2 Introduction

In the other chapters of this thesis the approach taken to study the effect of melanopsin has mirrored standard practice: we think that melanopsin might be involved in a specific process and so we run an experiment where we vary the amount of melanopsin activation within a stimulus (our independent variable) and measure something to see if there is an associated change in the target dependent variable, with the hope of understanding how melanopsin might be involved in the target process. However, at no point do we really drill down on the questions of why melanopsin might be involved in this process.

This has meant that it is unclear whether our results (positive or negative) can be taken at face value - perhaps our baseline assumptions about how or why melanopsin is involved were wrong, and we were consequently looking in the wrong place.

This chapter describes an exploratory computational study which took an ecological modelling approach to further our understanding of what, if any, benefits may arise by the use of a melanopsin-based signal for colour constancy. The proposed benefits of this approach are:

1. It may be possible to answer the question of whether it is sensible for melanopsin to be involved in colour constancy in any form whatsoever - is there any benefit to be gained from involving melanopsin?

2. We may be able to suggest or rule out specific computational structures - one computation might be beneficial whilst others might not be.
3. It may help to narrow the search area for future psychophysical experimentation - in previous experiments there have been lots of assumptions about the luminance range over which melanopsin is active, the spatial distribution (both in terms of receptive fields and influence over signals from distant parts of the retina), the temporal dynamics of the signals involved (and many other assumptions both conscious and unconscious). This is an opportunity to test whether a melanopsin-based signal is useful for specific ranges of the stimulus space.

The key research question for this section can be posed thus:

Considering the conditions on our planet, and the ecological requirements of vision, would a signal from a melanopsin-expressing cell be useful for colour constancy?

This chapter will describe the computational investigation and accompanying thought process in a roughly chronological order.

7.3 Data sources

7.3.1 Illuminant and Surface Data Sources

A useful guide to some of the existing datasets has been provided by Kohonen et al. [174], though many of the links have rotted since publication. Additionally, some new datasets have become available, and some datasets that weren’t included have become known to me. In this section I shall describe the datasets available for use in studies such as this.

7.3.1.1 Daylight datasets

It is standard practice (see for example Barrionuevo and Cao [16]) to use illuminants generated from CIE D-series formulae (see PTB function ‘GenerateCIEDay’) which are derived from data reported by Judd et al. [164]. Whilst the D-series provides a good approximation of daylight spectra, empirical data better represents any link between chromaticity and luminance, and any bias in the likelihood of one
spectrum occurring over another. It is thought that the original data of Judd et al. is no longer available [212, p. 60]. The first three principal components of the data are available through PTB as ‘B.sieday’.

**Granada Data.** The Granada daylight database [145] contains 2600 measurements of daylight taken over the course of two years at a single site in Granada, Spain. Data is recorded for 300-1100nm with a sampling interval of 5nm.³

**Other sources.** The Parkkinen and Silfsten data described by Kohonen et al. [174]⁴ comprises 14 measurements of daylight from afternoon and evening. The wavelength range is 390nm - 1070nm, with 4nm intervals.

The other potential sources of data, in addition to the Judd et al. [164] data, do not seem to be currently available, but for completeness I provide them here: Bui et al. [37], Condit and Grum [60], DiCarlo and Wandell [79], Dixon [80], Henderson and Hodgkiss [142], Sastri and Das [262, 263], Tarrant [293], Taylor and Kerr [294], Williams and Smith [321].

There are two authoritative reference books on the subject: Henderson [140, 141] (first and second editions) and [259]. Also of interest may be Minnaert [220] (various editions), and Lynch and Livingston [203] (various editions).

Two further datasets which are available only upon request are held by Dr Andrew Smedley of The University of Manchester (320nm to 2800nm, since 2010, data collection ongoing) and Marina Khazova of Public Health England⁵ (350nm - 830nm, 1nm interval). It is hoped that these datasets may be made openly available at some point in the future.

Data specifically for dawn and dusk (with a small amount of data extending into what could be considered ‘daylight’) is available from Spitschan et al. [282] as open access supplementary material from the journal publisher.

An interesting additional source of data may be the work of Peyvandi et al. [250], who simulate a very large number of daylight, sunlight and skylight spectra.

Finally, there is also a large corpus of information specifically about the light

---

³This data has been made available at http://colorimaginglab.ugr.es/pages/Data
⁴Available at http://cs.joensuu.fi/spectral/databases/download/daylight.htm
⁵Minimally described here: https://uk-air.defra.gov.uk/research/ozone-uv/uv-uk-monitoring
conditions in forest environments, although I have not had an opportunity to investigate whether collected datasets have been made available [19, 31, 44, 73, 94, 100, 101, 119, 126, 291, 296, 316, 324].

7.3.1.2 Surface Reflectance datasets

**Krinov data.** The Krinov data was originally published in 1947 [178], though it is now mainly accessed through a Canadian translation published a few years later [177]. It has recently been made available through PTB [23] (as sur_krinov.mat), and forms part of the SFU dataset [15]. It consists of 370 measurements of natural surfaces, measured at 9 locations around the USSR. It includes a large number of repeated measures (generally of objects at different angles), and has many measurements of objects which might be described as ‘background’ surfaces rather than objects per se (e.g. soil, sand, turf). Measurements are available at a sampling interval of 10nm, mostly between 400 and 650nm, with some extending as far as 900nm, and some without data at parts of the range. The PTB version of the data is a reduced set of 191 measurements, having excluded a number of measurements of various types of grass.

**‘Natural Colors’ data.** The ‘Natural Colors’ data [163, 244] was collected to allow investigators to explore how well reflectances could be represented by low dimensional models. The data consists of 219 reflectance spectra (though information provided by the authors suggests there should only be 218) of different leaves and flowers, between 400 and 700nm at a sampling interval of 5nm. It has recently been made available through PTB (as sur_koivisto.mat).

**Vrhel et al. data.** The Vrhel et al. [313] data in its complete form comprised measurements of 64 Munsell chips, 120 Du Pont paint chips and 170 natural and non-natural objects. Similarly to the ‘Natural Colors’ data, this data was again collected to allow investigations into the dimensionality of natural reflectance functions. The authors noted that they aimed to improve upon the Krinov data by decreasing the sampling interval (to 2nm), increasing the range of objects measured (and focusing on more object-like objects as opposed to background objects) and increasing the sampling

---

6It is available at: http://www.uef.fi/web/spectral/natural-colors
range (to 390-730nm). To my knowledge only part of this set is currently available, as the FTP server referenced in the original publication is no longer accessible. The object reflectances alone are available through PTB (as ‘sur.vrhel.mat’).

The Derby Set. One data-set which has been made available very recently is the data of Cheung [43]. The website from where they are now available\(^7\) states that ‘The reflectance factors of 274 objects (mainly leaves) were collected directly using a Macbeth 7000A reflectance spectrophotometer with specular component included. Each object was measured front and back to create 494 spectra.’ Unusually, photographs of each object measured are available from the same source. Cheung and Westland [42] describe this source as containing ‘leaves, petals, grasses and barks’.

Standard Object Colour Spectra Database for Colour Reproduction Evaluation (SOCS) data. This international standard [161, 292] collates more than 50,000 spectral reflectances of a wide range of type of surfaces, grouped into several categories. The database was originally created in order to allow for the assessment of colour reproduction of image input devices. Unfortunately, this data proves very difficult to access, and as yet I have been unable to assess it.

NASA data. The NASA data-set [70] comprises 156 measurements of different terrains and materials, presented to aid in the design of remote imaging systems to optimally detect surfaces of interest and to detect changes over time in these surfaces where this is of interest (e.g. changes in spectral signatures that reveal growth or disease of specific crops). Data was not collected by the authors, but digitised from 58 different sources, and so range and interval are not consistent throughout the set. Whilst the authors seem to have devoted a great deal of energy and care to accurate digitisation, ‘digitisation’ seems to be limited to the printing of tabulated values rather than provision of digital files (we’ve come a long way since 1985) and so any use of this data may need to start with an extended period of careful transcription. The surfaces chosen for this set are sensibly biased towards those of interest to remote sensing applications, and so use of this data in vision science would likely require careful consideration. It is expected that there may

\(^7\)http://stephenwestland.co.uk/spectra/index.htm
be other similar datasets tailored to the needs of remote sensing which may be available, should this type of data be appropriate.

Foster et al. hyperspectral images. The hyperspectral images of Nascimento et al. [232] and Foster et al. [114] provide nominal SRFs for full natural and suburban scenes. This data is valuable and rich in many ways.

Notably, it can begin to represent the ubiquity/rarity of certain types of reflectances in the natural world, whereas the statistical distribution of surface variability in the abstracted databases so far considered is at the mercy of the collator. As Maloney puts it: “in sampling spectral reflectances, we weight each spectral reflectance by its frequency of occurrence under whatever selection procedure we choose” [212].

Additionally, the spatial inter-relationships between surfaces can be considered, which may be of particular value in trying to understand how an organism might operate under real-world conditions.

Figure 7.1: A visualisation of the hyperspectral data for the first four images of the Nascimento et al. [232] data. The other four available images are of non-natural environments.

However, caution must be taken when using such data; whilst the hyperspectral images available are nominally ‘reflectance’ images, the way in which reflectance is computed may make them unsuitable for some uses. Reflectance is estimated from radiance images by assuming uniform illumination across the scene, which for some uses may be a particularly problematic simplification. This is an acceptably minor distinction for many use cases, but in this specific case this introduces error in precisely the place where it needs to be avoided. In considering the effectiveness of chromatic adaptation transforms the goal is to separate the effect of variable reflectance functions from variable power distributions, and the
ability to do this is hindered if an element of the power distribution variability is baked into the reflectance functions.

A final note here - whilst the use of spectral reflectance data from natural sources is often preferable to that from non-natural sources, it is possible that the careful use of non-natural data could be permitted following the finding of Maloney [209] that basis elements derived from measurements of Munsell colour samples provide excellent fits to natural data (specifically, the Krinov data).

Going further, it may be possible in some cases to use entirely artificial data; Chen et al. [40] showed that an artificial dataset, generated following the physical constraints on real SRFs (as discussed by Nassau [234]), seems to strongly resemble real datasets.

### 7.3.2 Chosen Data sources

Foundational data consisting of a subset\(^8\) of the Granada daylight dataset [145]\(^9\), the Stockman-Sharpe 10° cone fundamentals [288, 289] (aka the CIE 2006 10° cone fundamental sensitivity functions [51])\(^10\), the melanopsin fundamental of Lucas et al. [196]\(^11\), and a subset\(^12\) of the reflectances of Vrhel et al. [313]\(^13\) were used.

Tristimulus values \([L, M, S]\) and analogous melanopic values \(I\) were computed as per Equation 2.7 for each illuminant (as per standard tristumulus values but using the melanopsin fundamental, see Equation 7.1). A real-world version of this would be to measure a spectralon tile (or other uniformly reflective surface) under each daylight condition. Tristimulus values were then computed for each surface under each illuminant.

\[
I_{MB} = \sum_{\lambda} \phi(\lambda)\tilde{r}(\lambda)\Delta\lambda
\]

where \(\tilde{r}(\lambda)\) is the melanopsin fundamental of Lucas et al. [196].

---

\(^8\)Every 20th value (130 of total 2600), to reduce compute time, and increase legibility of plots.

\(^9\)Data: [http://colorimaginglab.ugr.es/pages/Data#doku_granada_daylight_spectral_database](http://colorimaginglab.ugr.es/pages/Data#doku_granada_daylight_spectral_database)

\(^10\)Available as ‘T_cones ss10’ in PTB.

\(^11\)Available as ‘T_melanopsin’ in PTB.

\(^12\)10 surfaces were chosen from the Vrhel dataset, with a roughly even representation of skin tones, fruit, and vegetable/greenery.

\(^13\)Available as ‘sur,vrhel’ in PTB.
MB chromaticity co-ordinates were calculated, for both illuminant alone and for each surface under each illuminant, as per Equation 2.8\textsuperscript{14}. The MB chromaticities are plotted in Figure 7.2.

Support for various other options was included:

- A range of CIE D series illuminants\textsuperscript{15}.
- Scenes 1-4 of the Nascimento et al. [232]\textsuperscript{16} hyperspectral reflectance data.
- Scenes 1-5 of the Foster et al. [114]\textsuperscript{17} hyperspectral reflectance data.

The CIE D-series illuminants allowed for a smaller and more controlled daylight dataset and the Nascimento/Foster et al. data allowed for a more realistic distribution of reflectances.

### 7.4 Recovering the chromaticity of daylight

The simplest computational method to achieve colour constancy is through normalisation of receptor signals by the hypothetical receptor signals for the illuminant. It is the implicit assumption in previous experiments that a melanopic signal could act as a cue to the chromaticity of the illuminant.

This is what Maloney [210] refers to as the ‘RGB heuristic’, and in some ways only delivers rough colour constancy, but it is nonetheless valuable. In most situations appearance of a scene under two different illuminants will be roughly relatable via the chromaticity of those illuminants.\textsuperscript{18}

Initially, two questions were proposed:

1. Considering only daylight spectra (excluding reflective surfaces), can a melanopic signal predict the chromaticity of daylight?

\textsuperscript{14}Note to assist in the reading of associated code: the normalising factors $k$ were applied during the Equation 2.8 rather than Equation 2.7.

\textsuperscript{15}Generated with the PTB function ‘GenerateCIEDay’.

\textsuperscript{16}Data: https://personalpages.manchester.ac.uk/staff/d.h.foster/Hyperspectral_images_of_natural_scenes_02.html

\textsuperscript{17}Data: https://personalpages.manchester.ac.uk/staff/d.h.foster/Hyperspectral_images_of_natural_scenes_04.html

\textsuperscript{18}However, it should be noted that there is no mathematical reason for this to be strictly true due to different colour rendering properties.
2. Now considering also object reflectances, can a melanopic signal predict the chromaticity of the daylight?

7.4.1 First-level signals

In this chapter the terminology of Barrionuevo and Cao [16] is employed; following their usage, ‘first-level signals’ are those which are direct photoreceptor catches (or analogues, such as XYZ tristimulus values), and ‘second-level signals’ are those computed by comparison of one signal with one or more other signals (such as xy or MB chromaticity values).

In order to visualise the relationship between the chromaticity of the illuminant and the resulting receptor catches, three-dimensional plots were made displaying $l_{MB}$ against $s_{MB}$ against [$L, M, S, I$] in turn. MB chromaticity space was chosen due
to its status as a physiologically based chromaticity diagram, following the logic that if biologically plausible mechanisms are sought, then computations in a space best representing the real space are preferable. The plot for $I$ is shown in Figure 7.3. Other plots closely resembled this one. It can be seen that although there exists some relationship between the chromaticity values of the illuminants and the $I$ values, it does not appear as though one could be used reliably to predict the other.

There is an almost bimodal relationship; all that can be gained from this relationship is the understanding that if the $I$ value is above a certain threshold, it is likely to be in a group with a relatively low $s_{MB}$ and relatively high $l_{MB}$ value. The real-world correlate of this is - if the daylight is bright enough, I can be fairly confident that the chromaticity of light will be relatively warm in colour. This is as expected, since the brightest daylight conditions are likely to be those with unobstructed direct sunlight, which is warmer in CCT than illumination provided by the blue sky. This relationship could not be used to predict the precise chromaticity of an illuminant. A near-identical trend is seen when surfaces are considered (the coloured points in Figure 7.3).

### 7.4.2 Second-level signals

Following this, similar plots were made which plotted $l_{MB}$ against $s_{MB}$ as before, but now plotted the various second-level combinations of $[L, M, S, I]$, created by considering one signal divided by another (e.g. $L/M$), on the z-axis. These plots are shown in Figure 7.4. The plots created all showed clear and relatively simple relationships between chromaticity and these new derived signals, at least for the illuminant-only values.

A similar trend was seen for each individual surface as was seen for illuminant-only, however these relationships appear to be offset from that of a perfect reflector.

### 7.4.3 A PCA interpretation

These results seem to make intuitive sense if we consider the daylight dataset from a PCA perspective (Figure 7.5). For this dataset the first principal component (PC1) is broad and relatively smooth, and accounts for 99.888% of the variance, and so
Figure 7.3: MB chromaticity of 130 illuminants plotting against the $I$ values for the illuminants alone (black points) and 10 different surfaces (coloured points). The edge of the spectral locus is visible along the $l_{MB}$ edge of the diagram.
7.4. Recovering the chromaticity of daylight

Figure 7.4: Relationship between chromaticity and second-level signals. Plotted on the x-axis here is $l_{MB}$, with second-level signals on the apparent y-axis. During analysis these plots were three dimensional, with the apparent y-axis being a z-axis and the x-axis joined by a y-axis of $s_{MB}$. As before, black points indicate direct illuminant values, with coloured values corresponding to surface reflectances (though the colours do not relate the surfaces other than to distinguish one from another). It can be seen that on a per object basis there is good correlation between chromaticity and all of the above signals (mean correlation coefficient of 0.9782, SD of correlation coefficients of 0.0144, calculated for surface data only). A similar relationship holds for the $s_{MB}$ perspective (0.9821, 0.0073 respectively).
a first-level signal in any part of the spectrum is going to essentially track this component.

![Graph](image.png)

**Figure 7.5:** The first 3 principal components for the Granada daylight dataset. These first three components account for 99.888%, 0.077% and 0.015% of variance respectively.

The second principal component (PC2) is relatively smooth and monotonic (and accounts for 0.077% of the variance). Almost any comparison between signals at different points in the spectrum (second-level signal) is going to roughly track this component (though the relative contribution of this component will slightly depend on the value of PC1).

Each surface will reflect different parts of the spectrum, meaning that this sampling of PC2 will vary from surface to surface, but the same overall trend will be followed.
7.4.4 Relationships as artefacts

For many of these signals however, such relationships are likely to arise solely from the manner in which each signal is constructed\(^{19}\). For example, a relationship between \(l_{\text{MB}}\) and \(L/M\) might be expected, since \(l_{\text{MB}}\) is defined as \(L\) divided by the sum of weighted components of \(L\) and \(M\) (Equation 2.8). To confirm this, a control condition was performed, where \([L, M, S, I]\) was replaced by randomly generated values, and the second-level signals were generated as before. The results can be seen in Figure 7.6. Relationships between secondary signals derived from \(L\), \(M\) or \(S\) signals still showed correlations with chromaticity (albeit now points fell on a plane, instead of a line), whereas signals with an \(I\) component showed only minimal coherence, forming a noisy cloud in three-dimensional space.

The fact that the relationship between chromaticity and second-level signals degrades for a melanopsin-based second-level signal, but not for a cone-based second-level signal, when data is replaced with randomly generated noise, suggests that whilst the relationships in the cone-based cases may arise simply due to the mathematical similarity between the calculation of chromaticity and second-level signals, this is not the case for the melanopsin-based second-level signals.

Considering that relationships between the melanopsin-based second-level signals and chromaticity do seem to exist (Figure 7.4), one must conclude that there is a regularity in the data which allows for such a relationship to be revealed.

To summarise this section:

1. A basic melanopic signal, computed from direct daylight measurements, cannot predict chromaticity (even in a one-dimensional sense, such as predicting CCT), other than a crude estimation of whether a measurement is direct sunlight or not. Figure 7.3

2. The same can be said for melanopic values computed for daylight reflected off a surface. Figure 7.3

\(^{19}\)My thanks go to Manuel Spitschan for convincing me of this point.
Figure 7.6: As per 7.4 but where \([L, M, S, I]\) was replaced by randomly generated values, and the second-level signals were generated from these instead of real values. Blue is used to distinguish this random data from previous real data. Note that different rotations have been applied to some of the subplots to best display the correlations.
3. There are relatively strong and simple relationships between hypothetical second-level signals and illuminant chromaticity. Figure 7.4

4. When considering surfaces, these relationships are maintained, but offset differently for each surface. Figure 7.4

5. Some of these correlations appear to be nothing more than computational artefacts (Figure 7.6). Notably, the melanopic second-level signals did not seem to be such artefacts.

7.5 Transformation to an illuminant-independent space

Whilst estimating the chromaticity of the illuminant is one way in which colour constancy might be achieved, it is also possible that some transformation might exist which delivers signals into an illuminant-independent space but which doesn’t explicitly depend upon an estimate of the chromaticity of the illuminant.

Considering that the absolute melanopic values seemed to offer little predictive benefit (see Figure 7.3) the choice was made to create a normalised melanopic value similar in nature to the MB values. Following Equation 2.8, but with a melanopic fundamental (as per Equation 7.1), allowed for the creation of what will be termed the ‘$i_{MB}$’ values.

\[
i_{MB} = \frac{I_{MB}}{L_{MB} + M_{MB}}
\]  

(7.2)

where $I_{MB}$ follows from Equation 7.1, and $L_{MB}$ and $M_{MB}$ follow from Equation 2.7.

7.5.1 A melanopic signal as a third dimension

In order to understand the relationships between the MB chromaticities and the new $i_{MB}$ values, and to understand what types of corrective computations might be effective, consideration was given to the $i_{MB}$ signal as a third dimension upon the already two-dimensional MB chromaticity space, such as shown in Figure
7.7 (akin to Figure 7.3, but for this new second-level signal). This allowed for the consideration of what properties a signal in this third dimension would need to possess in order to be able to transform the values into a two-dimensional illuminant-independent space.

It was found that a log transform applied to the $[I_s i]_{MB}$ data made relationships between the various signals more linear, and the data much more normally distributed. A power transformation was also effective in this goal, but a log transform was chosen such that multiplicative transformations could be implemented as additive transformations at later stages. The data was also zero-meaned and normalised for standard deviation so that the signals were on like scales.

A range of speculative transformations were performed, to transform colour signals to illuminant-independent colour signals. A rotational transformation was successful (analogous to changing the perspective as in figure 7.8) and additionally a weighted additive model was found to be rather successful. No satisfactory model
was found where a purely multiplicative transformation was applied.

Thinking about a corrective signal as an additional dimension upon a MB chromaticity space allows for consideration of requisite or desirable properties that this third signal should imbue upon the three-dimensional cloud of points. I have identified one requisite property, and one desirable property.

The requisite property for such a signal, as already suggested, is that it should differ from other chromatic signals enough to allow for the point cloud of chromatic points to be significantly non-planar. This in turn allows for a projection upon two-dimensional space which has the potential to be illuminant-invariant.

The desirable property is that it should be roughly monotonic with respect to other chromatic signals. This allows a one-to-one mapping of colour signals to illuminant-independent colour signals, where a non-monotonic relationship does not necessarily do so. This is not a strict requirement, since the corrective function only needs to be unidirectional, but non-monotonicity would exclude simple transforms (such as linear additive or rotational).

Another key finding at this stage was that there is a perspective upon a three-dimensional point cloud \([l_{MB}, s_{MB}, i_{MB}]\) from which points from like objects clustered well, such as in figure 7.8. This property shows that there is at least one transformation such that the points could be projected onto a two-dimensional plane where the illuminant-dependence would be greatly reduced whilst the inter-object chromatic relationships were retained.

Note that such a perspective would not be possible for a space where the third dimension was a near-duplicate of the first or second, since in this case all points would lie upon a plane. This would be the case in circumstances where a corrective signal too closely resembled the visual signals. The fact that points do not lie on a plane shows that there is some decorrelation between \(i_{MB}\) and both \(l_{MB}\) and \(s_{MB}\). The fact that they do not lie on a plane and instead diverge from this plane in a surface-dependent fashion (as opposed to noisily diverging) suggests that the \(i_{MB}\) may present a means to separate the contributions from illuminant and surface.
Figure 7.8: Figure 7.7 rotated to a perspective where it can be seen that the points project upon a two-dimensional plane in a clustered fashion whilst not degrading to a line within that space. A physiological analogue of such a rotation would be a recombination of weighted signals.

7.5.2 Additive/Subtractive model

It was found that by addition or subtraction of weighted \( i_{MB} \) values, as per equation 7.3 conversion to an approximately illuminant-independent space could be performed. An optimisation initially sought optimal weights for the scaling factors with the goal of minimising the overall spread of chromaticities (measured as standard deviation of the entire set). This was later amended to the intra-group spread of chromaticities, measured as standard deviation within each group. This had a relatively minor impact, resulting in coefficient values of \( k_1 = 0.57 \), and \( k_2 = -0.94 \).

\[
\begin{align*}
l_{MB}^* &= l_{MB} + k_1 i_{MB} \\
s_{MB}^* &= s_{MB} + k_2 i_{MB}
\end{align*}
\]
7.5. Transformation to an illuminant-independent space

where following the optimisation \( k_1 = 0.44 \), and \( k_2 = -0.91 \).

The results of applying Equation 7.3 can be seen in figure 7.9. The effect of different weights can be seen in Figure 7.10. It can be seen that the points now roughly cluster together by object, with less smearing induced by changes in illuminant (compared to Figure 7.2).

![Figure 7.9: MB chromaticities for 10 reflectances under 130 daylight illuminants, corrected by the corresponding \( i_{MB} \) value as per Equation 7.3.](image)

It was unclear however whether this advantage would be gained by the addition of any new signal, or whether the melanopsin signal was in any way optimal. With this in mind, the peak spectral sensitivity of the melanopsin function was parameterized, to consider whether shifting it along the wavelength spectrum affected the performance of such a signal. Each new sensitivity function was allowed the benefit of its own weighting optimisation, allowing different scaling factor weights. The results of this computation can be seen in figure 7.11.

The standard deviation minima for \( l_{MB}^* \) occur at 532nm and 590nm, and for \( s_{MB}^* \) at 444nm and 568nm. This would suggest that the optimal spectral sensitivity
for a receptor employed to correct the signals from the various cones would be at these points in the spectrum. However, if we plot the transformed chromaticities computed using hypothetical signals at these wavelengths, we see that the results are not an obvious improvement. In Figure 7.12 the results of using a nominal melanopic signal with a peak spectral sensitivity of 590nm are shown. Group SD is low, but different surfaces are not particularly well distinguished from each other. We witness good colour constancy, at the expense of chromatic discrimination\textsuperscript{20}. This is obviously not ideal.

This result made it very clear that more thought was required regarding how ‘success’ was calculated.

The actual desired behaviour would be to minimise internal variance for each reflectance group (rather than the whole set), whilst maintaining at least some

\textsuperscript{20} At my poster at VSS David Brainard referred to this as the 'Ford Model of Colour Constancy' - you can have any colour, so long as it’s black.
To summarise this section:

1. Following earlier findings that a second-level signal seemed more likely to show a relationship to chromaticity, and be likely to sample the PC2 values of daylight, hypothetical $i_{MB}$ values were computed.

2. Various transformations were trialled: additive/subtractive, rotational and multiplicative. The first two showed promise (and seemed to be roughly interchangeable), whereas no adequate multiplicative transformation was found.

3. One requisite and one desirable property for a hypothetical third signal were identified: decorrelation (leading to non-planarity) and monotonicity.
4. A perspective upon the three-dimensional $[Lsi]_\text{MB}$ space was found where points diverged from a plane in a surface-dependent fashion. *Figure 7.8*

5. A basic additive/subtractive algorithm was applied, which seemed to reduce the spread of chromaticities in a meaningful fashion. *Figure 7.9*

6. It was found that when the melanopic value was replaced by signals generated from different spectral sensitivities, some hypothetical signals appeared to be more successful than a melanopic value. These signals performed very well by the metric of whole-set standard deviation, but in practice would make very poor corrective signals due to the reduced discriminability between surfaces. *Figure 7.12*
7.6 Assessment of colour constancy algorithms

The basic mechanism of an additive/subtractive model of melanopic colour constancy seems worthy of further scrutiny. However, we have seen that care is required in defining what is considered successful performance for a colour constancy algorithm. Broadly speaking, we would like our algorithm to be judged on its ability to allow for recognition (reliably remapping points from a specific surface to a specific point in appearance space) and discrimination (creating distinct representations of different surfaces).

Different schools of thought within colour constancy research have used different methods for assessing proposed algorithms, mirroring subtle differences in their respective conceptual goals. Three subtle variants can be described via representative research questions:

- How do humans achieve colour constancy?
- How do we predict what humans perceive?
- How can we colour correct digital photos to emulate the original appearance of the scene to the photographer?

The first group have tended to assess their proposed solutions in the context of biological and physical plausibility. Where solutions in this group rely upon some regularity in the environment, discussions gravitate to the value, ubiquity, and veracity/validity of this regularity (see Hurlbert’s table [156, p. 295] which sorts a range of colour constancy algorithms by the assumptions that they rely upon). Alternatively, studies can remove the cue or cues under investigation from a controlled visual scene and in some way measure the impact upon an observer’s state of chromatic adaptation [176]. I am not aware of a standard measure for quantifying the ability or biological plausibility of a particular proposed algorithm.

\[\text{The ideal colour constancy algorithm would distinguish between all spectrally distinct surfaces, no matter how minor their distinctions. However, here we are more interested in uncovering a mechanism which may be used by the human visual system, and so it seems sensible to limit the requirement to separation of surfaces which would generally be seen as chromatically distinct.}\]
The second group have made considerable progress in the development of the CATs which are used within CAMs, through the collection of corresponding colour datasets (pairs of colorimetric values where the first member of each pair under a first illuminant matches in appearance the second under a second illuminant). Models are then fitted to these datasets such that the appearance of an arbitrary colour under the one illuminant could be predicted under the other, and vice versa. The effectiveness of a model is quantified by considering how well the model fits data that was independently collected [57].

The third group focuses on estimating the tristimulus values or colorimetry of an unknown illuminant from the pixels of a digital photographic image. There is obvious crossover between this group and the previous groups, for two reasons. Firstly, the human visual system could compute relatively meaningful estimates of surface reflectance properties if the illumination spectra could be estimated with moderate accuracy [212]. Secondly, where the goal is to output an image which appears as it would to a human observer, there is implicitly a goal to understand how an observer might be performing this computation. This group traditionally quantifies success of algorithms by how well they can compute an illuminant estimate in terms of its RGB values in a camera-specific colour space.

There is no clear fit between existing models of assessment and the specific goals of the algorithm under assessment. The most natural fit is within the first group, since the key interest is whether a melanopsin-based signal provides a valuable tool to a biological system in solving an ecological problem (and thus whether it might exist in real biological systems). However, it is unclear at this stage what is the mechanism by which such a system may operate, and so the discussion of what assumptions such a mechanism may rely on seem overly abstract.

Of clearer applicability is the work of Barnard et al. [13] (later used by Hordley and Finlayson [147] and Gijsenij et al. [127]). They propose a method whereby spectral measurements of illuminants and surfaces are used to create ‘synthesized’ data, representing each surface under each illuminant, which can be used as an input to a colour constancy algorithm. Since the ground-truth is known here, the
results can be quantitively assessed.

There are some amendments to their process which are required if we are to use it here. Firstly, the data used in their experiments comprises natural and non-natural surface reflectances and illuminants, which is suitable for their purposes of assessing algorithms designed for digital cameras, but is unsuitable here for assessing an algorithm to understand its ability in a natural environment (the environment which would have influenced the natural evolution of biological systems). Therefore, the reflectance and illuminant data will need to be limited to naturally occurring data. Secondly, their performance metrics again relate to their problem rather than ours: they consider only the accuracy of the estimated white point, and whilst this may be useful to a biological system, the grander goal is a conversion of chromaticities to an illuminant-independent surface appearance space. An illuminant estimate is one way to reach this goal, but it is not fundamentally necessary, and the accuracy of this value does not necessarily provide a good measure of effectiveness of the whole system.

One performance metric which more closely matches the goals of this study is that provided by Forsyth [112, p. 19], who suggests using “the median Euclidean distance of the outputs from the average (over the different lights) output, for each [colour] chip. This median is then normalised by the Euclidean magnitude of the outputs.” This value provides an insight into the nature of the clustering of points following a transform, which can be considered as loosely related to our fundamental goal of allowing for recognition (a point being mapped to roughly the point it should be). However, no regard is given to our second fundamental goal of discrimination. Indeed, an algorithm which outputs two clusters of points which lay directly atop each other, even if these clusters represented vastly different objects, would be rated as successful, if these clusters were particularly tightly clustered.

### 7.6.1 K-means clustering as an assessment tool

Whilst the clustering metric of Forsyth is a step in the right direction, it doesn’t quantify discriminability, and there is no clear way in which it could be amended to do so. Further, it would be preferable to consider these two characteristics in
unison since they are valuable only in relation to each other: we could allow for loose clustering if the cluster centres were distant, but require tight clusters if they were close.

To address this shortcoming I propose the following: that the output of an algorithm (where the input is a batch of synthetic data such as in the method of Barnard et al. [13]) should be subjected to a k-means clustering (this is easy where synthesised data is used and $k$ is known) and this output in turn should be marked against the known groupings of the original input data. This marking would reveal how well the algorithm had remapped points into clusters relating to their original grouping, and how separable the clusters are. A suitable number of repetitions of the k-means clustering is advisable, to avoid local minima.

A computational demonstration of this procedure is shown in Listing 7.1, in rough MATLAB code\textsuperscript{22}. For each cluster (based on the ground-truth data), the proportion of points which are given the modal value (per ground-truth cluster), is computed. This value is then averaged over clusters. A perfect score, where each point within a ground-truth cluster was given the same value by the k-means algorithm, would be 1, with the chance baseline depending on the number of points and number of clusters.

\begin{verbatim}
GT = GroundTruth; % The actual cluster information
KMI = KMeansIndex; % Result from K-means clustering

for ii = 1:numberOfClusters
    MP(ii) = mean(mode(KMI(GT == ii)) == KMI(GT == ii)); %mode proportion
end

KMM = mean(MP); % k-means-mark
\end{verbatim}

To test/demonstrate such a procedure, let us consider the plots from Figure 7.9 and 7.12. Whilst the simple measure of whole-set standard deviation reports the

\textsuperscript{22}A runnable version of Listing 7.1 is available (https://github.com/da5nsy/Melanopsin_Computational/blob/master/KMeansMarkDemo_abstract.m).
latter as the more successful output, the k-means-mark for these outputs is 0.8915 and 0.5662 respectively, which tallies with our intuition of which instance of the algorithm has performed better. The K-means selections for each are visualised in Figures 7.13 and 7.14.

Before the k-means clustering algorithm was applied, data were again zero-meaned (no effect on algorithm, only neatens later presentation) and normalised by standard deviation (medium/large effect on algorithm, resulting in equal weights given to variance in either dimension of MB space).

![Figure 7.13](image)

**Figure 7.13:** The data of Figure 7.9, plotted with colours indicating the groupings chosen by a k-means clustering procedure. Note that the colours do not relate to appearance of surfaces, they are only to indicate groupings. Note also that these colours do not correspond to colours used in the previous figure; the k-means-mark algorithm is not interested in the absolute labelling of groups, rather whether all members from group X have been sorted into group Y etc.)

### 7.6.2 A comparison of four algorithms

The simple melanopic algorithm (additive/subtractive) was compared with 3 other algorithms: a GW algorithm, a BiW algorithm and a ‘Do Nothing’ (DN) algorithm (a baseline condition where no correction was applied). The structure of Barnard et al. [13] was employed, but with the modifications suggested above.
Figure 7.14: The data of Figure 7.12, plotted with colours indicating the groupings chosen by a k-means clustering procedure.

The results of the four algorithms are shown in Figure 7.15. It can be seen that the GW algorithm performs particularly well under this test condition, separating the different surfaces into very tightly clustered and very distant groups. The BiW algorithm appears to choose one of two surfaces as the brightest, which splits each surface group into two groups. If it were able to pick a single reference point it appears as though it would be similarly successful. The melanopsin-based correction performs moderately well; as shown before it improves greatly upon the DN condition, but the groups are not particularly tightly clustered and there is a moderate amount of overlap between the groups.

This visual analysis is mirrored by the k-means-marks computed for each output. They were: DN: 0.5869, GW: 0.9631, BiW: 0.5308, mel-based: 0.9346.\textsuperscript{23}

### 7.6.3 A more naturalistic test case

In terms of real-world validity, this test condition represents a rather unusual situation: a set of 10 surfaces is represented under each illuminant, with each

\textsuperscript{23}It should be noted that due to the random nature of initial seeding in k-means clustering, these values are expected to vary slightly depending on initial state of the random number generator.
Figure 7.15: The output of the four algorithms; top left: DN, top right: GW, bottom left: BiW, bottom right: melanopsin-based correction, as per Equation 7.3. Colours denote different surface groups.
Figure 7.16: As per Figure 7.15 but where only 50% of the surfaces are present under each illuminant.
surface always being present, in the same proportion (relative to other surfaces) each time. Under these conditions, we should expect GW and BiW to flourish. In more realistic conditions however (where the scene was less stable), we might expect them to suffer.

When we randomly remove 50% of the surfaces under each illuminant, the resulting k-means-marks are: DN: 0.5989, GW: 0.6967, BiW: 0.5916, mel-based: 0.9447. It can be seen that GW has fallen dramatically whilst the others have remained roughly stable.

A visualisation of the results at under this condition is shown in Figure 7.16. It can be seen that the neat clusters that GW showed in Figure 7.15 have degenerated into broad clouds. This can be explained by the cue that GW uses becoming noisier; when viewing unknown surfaces drawn from a normally distributed set of surfaces (in terms of chromatic bias), the fewer surfaces that are visible, the lower the likelihood that the mean chromaticity of the surfaces represents the chromaticity of the illuminant.

Parameterising the percentage of surfaces shows that this general trend continues at other percentages. Outcomes at percentages from 20% to 100% in 10% intervals are shown in Figure 7.17.

Increasing the number of surfaces has little effect on the overall trend. A trial performed with 32 surfaces from the Vrhel et al. [313] dataset (selected such that they were ‘natural’ and above a minimum threshold chromaticity difference when chromaticity was calculated under a single daylight measurement) showed a slight overall decrease in the performance of the melanopsin-based correction, but it still outperformed all other algorithms below 80% surface inclusion. Interestingly, the BiW algorithm performed much better than it had in previous trials, presumably within this subset of the data the brightest surface was a more reliable cue than in the previous set. A summary for this test is shown in Figure 7.18.

To summarise this section:

1. The analysis of colour constancy algorithms by previous research groups was considered, noting their distinct aims within the colour constancy field.
Figure 7.17: The performance of the various algorithms as a function of the available surfaces (100% = 10).

(a) The framework of Barnard et al. [13] was employed: synthetic stimuli were produced.

(b) The clustering method of Forsyth [112] was considered, but found unsuitable.

(c) A modification whereby the algorithm output was subjected to a k-means clustering analysis, and the computed clusters were compared to ground-truth clusters, was proposed.

2. Four colour constancy algorithms were tested.

(a) Under the baseline conditions, whereby each surface was presented once under each illuminant, the GW algorithm performed best, followed closely by the melanopsin-based transform.

(b) Under more naturalistic conditions, where a new random subset of 50% of the reflectances was shown under each illuminant, the performance
7.7. Optimality of the melanopsin spectral sensitivity

Now that we have a seemingly robust method for quantifying the performance of an algorithm, we are better placed to consider whether the SSF of melanopsin is in any way ‘optimal’ for this task, or whether any performance stems simply from allowing this particular algorithm to have access to a greater amount of information through additional sampling of the scene.

Re-running these computations but shifting the spectral sensitivity of the nominal melanopsin SSF to lower and higher wavelengths, gives Figure 7.19. At each wavelength a new optimal pair of $k$ values (Equation 7.3) was computed. The computations were re-run 10 times, and the minimum, mean and maximum

Figure 7.18: As per Figure 7.17 but with an extended set of 32 SRFs.

of the GW algorithm dropped drastically.

(c) This trend was shown to hold at other percentages of surfaces included and for larger surface sets.

7.7 Optimality of the melanopsin spectral sensitivity

Now that we have a seemingly robust method for quantifying the performance of an algorithm, we are better placed to consider whether the SSF of melanopsin is in any way ‘optimal’ for this task, or whether any performance stems simply from allowing this particular algorithm to have access to a greater amount of information through additional sampling of the scene.

Re-running these computations but shifting the spectral sensitivity of the nominal melanopsin SSF to lower and higher wavelengths, gives Figure 7.19. At each wavelength a new optimal pair of $k$ values (Equation 7.3) was computed. The computations were re-run 10 times, and the minimum, mean and maximum
taken for each point. This variability exists only because of the random seeding required for the k-means clustering algorithm. Within the standard script there are 20 repeats of each k-means-mark test, and presumably if this figure was raised the variability visible in Figure 7.19 would disappear.

![Graph](image_url)

**Figure 7.19:** The results of varying the peak wavelength of the nominal melanopsin spectral sensitivity function.

It can be seen that the baseline spectral sensitivity of melanopsin (peak at 488nm) is in the best area of the spectrum for a signal which is to be used for correction in this fashion (peaking at around 510nm). There is an additional peak at around 582nm. It may be coincidental, but it seems worth noting that the spectral sensitivity of the bistable form of melanopsin would be at 587nm [229].

### 7.8 Why is this possible?
At this stage the underlying reason for the relative success of these transforms was considered. One potential lead is a curious regularity at the shorter wavelengths of the reflectance spectrum of many natural objects. In figure 7.20 it can be seen
that there is a plateau in the relative spectral reflectances of some natural objects (a subset of the Vrhel et al. [313] reflectances) between approximately 430nm and 480nm, with relatively small deviations in the range 400nm to 500nm. Another way to visualise this is to calculate the correlation between points on the reflectance function. Such a calculation was performed initially on a combined set of scenes 1-4 of the Nascimento et al. [232] hyperspectral reflectance data and scenes 1-5 of the Foster et al. [114] hyperspectral reflectance data, and then on three other datasets. See Figures 7.21 and 7.22 respectively.

![Figure 7.20: The SRFs of a subset of the Vrhel et al. [313] reflectances.](https://personalpages.manchester.ac.uk/staff/d.h.foster/Hyperspectral_images_of_natural_scenes_02.html)

It may be possible to take advantage of this regularity. Assume a simpler situation where there were two points on the spectrum of the reflectance functions of a set of natural objects which were always the same as each other (perfect correlation). Sensors of appropriately narrow spectral sensitivity, placed at these two

---

24Data: https://personalpages.manchester.ac.uk/staff/d.h.foster/Hyperspectral_images_of_natural_scenes_02.html
25Data: https://personalpages.manchester.ac.uk/staff/d.h.foster/Hyperspectral_images_of_natural_scenes_04.html
Figure 7.21: Visualisation of the average correlation matrix for the scenes 1-4 of Nascimento et al. [232] and 1-5 of Foster et al. [114]. Note the lighter square region in the top left, indicating an area of increased correlation across different wavelength samplings, and another square region (though with a faded lower-right corner) in the centre. The dark frame to this figure is likely due to increased measurement noise at these extremes. The figure is symmetric about the diagonal axis.

points on the spectrum would always register corresponding signals. Considering that the second principal component of daylight variability is a broad skew, any difference in the signals from two such receptors could be used fairly reliably to sense the contribution of this second component in any single condition.

7.9 Physiological plausibility

There are multiple ways in which a transformation such as (or similar to) the one investigated here could be implemented in a biological setting. As discussed at the end of Section 2.2.4, there are a number of sites at various stages of the visual system which could support such a transformation. Perhaps the simplest would be for ipRGCs to modify their standard (cone and rod based) outputs as a function of
melanopic activation, creating a subset of nerves which provided ‘colour constant’ perception from this early stage of the visual process. It also seems plausible that intraretinal networks could be employed to modify a greater range of signals before transmission by traditional RGCs. It is also possible that a melanopic signal could be used much later in the visual system - perhaps even to the extent that it could occur as a percept [284] that helps us somewhat consciously achieve colour constancy, far beyond the realm of what one may consider chromatic adaptation.
7.10 Implications for future research

One aim of this exploratory study was to narrow the search space for future investigations, and on this topic I have a number of suggestions, based on both this study and the experiments of previous chapters. I shall attempt to focus on the suggestions which arise directly from the results presented in this chapter.

The target variable in many previous studies has been melanopic activation. No clear route has been shown here which would indicate a mechanism by which a raw melanopic signal could be used, to the extent that a melanopic signal cannot predict the chromaticity of daylight. A more plausible candidate at this stage is a normalised melanopsin value of some sort. The practical implication of this is that the need to control luminance and other cone/rod signals is paramount. If it transpires that a raw melanopsin signal is used, then there is no harm done by following this suggestion.

In order to assess whether melanopsin is involved in colour constancy at a full system level, a more holistic test than achromatic matching is required. Colour naming or discrimination of illuminant change from reflectance change seem clear candidates, as both use semi-natural tasks that, under the best circumstances, would require an observer to use the full extent of their visual system.

It is possible that the use of naturalistic contrasts may be important. For example, assuming there are multiple cues which are used to enable colour constancy, even if a melanopic process is one, it is possible that it may be ‘overruled’ if multiple other mechanisms suggest a different conclusion to that which would be gained from using the melanopic process alone. To be clear - it seems possible that there might be a naturalistic sweet-spot, too little melanopic activation clearly won’t activate any melanopic processes, but it is possible that too much melanopic activation, relative to other visual signals, may also result in a non-representative finding. Allen et al. [4] took a step in this direction by using the hyperspectral images of Foster et al., discussed in Section 7.3.1.2, to understand the levels of melanopic contrast which exist in real scenes.
7.10. Implications for future research

Considering the available daylight datasets\textsuperscript{26} there appears to be a single primary axis of chromatic variation. It is to be expected, based upon ecological requirements but also our own experience of the world (where constancy seems effective), that constancy would be best suited to this type of variation. All other types of adaptation would seem substantially less important, if judged on the magnitude of the effect they hope to nullify alone\textsuperscript{27}. Thus it seems wise to prioritise testing of adaptation along this specific vector, with other vectors being of secondary concern.

\textsuperscript{26}I suspect that existing daylight datasets downplay the importance of the variability of inter-reflections from surroundings. Specifically, when moving through forest environments I suspect that there is a great deal more variation in a red/green direction than would be suggested from daylight datasets taken from a single unshaded position.

\textsuperscript{27}It is of course possible that there are specific surfaces for which constancy is vitally important, and that these specific surfaces may vary along subtly different vectors than the primary vector along the caerulean line.


7.11 Conclusions

In this exploratory computational study, the goal was to understand whether a signal from a melanopsin-expressing cell might be useful for colour constancy.

The first question was whether a melanopic signal could be used as a cue to the illuminant. Following the investigation, this seems like rather a naive question, but it is worth recalling that this (or something close) is the underlying assumption in our previous studies. It is now clear that a raw melanopic signal (not normalised in any fashion) can only provide a very crude cue to the chromaticity of the illuminant: if it is above a certain threshold then it can be assumed that the illuminant on the scene is direct sunlight, which has a relatively reliable chromaticity. This signal notably is also strongly correlated with the other retinal signals, and therefore provides no clear additional value.

One line of enquiry which may be valuable to pursue, and which may provide a reprise for a raw melanopic signal, would be to consider the problem through the lens of a Maloney-Wandell-type linear models approach [211, 212]. This line of reasoning approximates both SPDs and SRFs as linear combinations of a small number of basis functions, and deduces several core rules regarding the relative number of bases, receptor classes, and surfaces in a single scene that would be needed to perform colour constancy (more formally - to reconstruct the SRF). The key finding was that to enable reconstruction of surface reflectance with three degrees of freedom a model would require four sensor classes. Maloney’s doctoral thesis concludes: “Whether human vision achieves the same endpoint with only three classes remains to be determined.” Written in 1984, ipRGCs were not yet known, and it was assumed that rods were not operational at high enough luminances to contribute meaningfully to colour constancy. It is expected that considering melanopsin within this framework would prove a very valuable endeavour. It is mentioned here since the input to such a system would be a ‘raw’ signal rather than a normalised signal.

The normalised, or ‘second-level’, signals tested here showed moderate promise. All possible combinations of cone signals and melanopic signals were
tested, and each was shown to be highly correlated with the chromaticity (in MB space) of the illuminant. Relationships between signals from surfaces and the illuminant chromaticity were similar on a surface-by-surface basis but showed considerable offset (Figure 7.4). It was shown that for the purely cone-based second-level signals these correlations could arise purely as computational artefacts based on how chromaticity is calculated, but that in the case of the melanopic signals no such inherent relationship existed.

Considering a melanopic signal as a third dimension upon a MB chromaticity diagram yielded two abstract requirements for a signal for which the purpose was to enable a transformation to an illuminant-independent space. The first was a moderate level of decorrelation with existing signals such that the cloud of points within this space was non-planar. The second was that this new signal should track an approximately monotonic path through this space. A three-dimensional perspective also afforded the view that there was at least one type of transformation which might use the $i_{MB}$ values constructively (Figure 7.8).

A specific transform was tested against a number of standard transforms, using a novel quantitative measure of success (‘k-means-mark’). By this measure, the GW algorithm was found to be the most effective for our basic condition. However, when the number of surfaces in a scene was randomly reduced the performance of the GW algorithm began to fall rapidly, whereas the melanopic algorithm was unaffected (See Figures 7.16 and 7.17). A melanopic algorithm does not rely upon stability of scene-level attributes, whereas GW and BiW do.

The optimality of the spectral sensitivity of melanopsin for providing a signal for such an algorithm was queried. The root of the question here was whether this is something which can be performed with any additional signal, or whether there was something special about the spectral sensitivity of melanopsin. Apart from the inherent interest of understanding whether there is an optimal sensitivity for such a signal, and what this optimal sensitivity is, there is the additional fact that if it could be shown that a melanopsin-based signal was optimal in some respects, in would lend credence to the theory that melanopsin is involved in this task, since
we assume that in the course of evolution the sensitivity may have shifted to a position of greatest value over time.

It was found that the SSF of melanopsin is in a spectral window where such a transform would be particularly effective (for our choices of illuminants, surfaces and cone sensitivities). However, the non-optimal signals also performed remarkably well, with many parts of the spectrum allowing for k-means-marks of above 0.8 (Figure 7.19).

The results of this work suggest that there is a mechanism by which an additional signal could be used fruitfully in the pursuit of colour constancy, and that a melanopsin based signal would not be a poor candidate for such a role.

7.12 Further work

There are a number of suggestions for further research:

- Further investigate why a melanopic transform might work.
  - Surface spectral regularities? Test with artificial surfaces and illuminants - one would expect the ability to diminish with non-natural surfaces and illuminants, particularly ones that did not conform to naturalistic constraints. See MacDonald [207, p. 239-40] for an example of such a set of synthetic SRFs.

- Simplify the grading mechanism.
  - Whilst the k-means-mark method provides a suitable method for determining functional equivalence between tightness of clustering and separation between clusters, it has a couple of annoying issues. Firstly, it takes rather a long time to compute (though I must remain thankful for the relative speed of modern computers) and secondly, it introduces a level of inherent uncertainty (as can be seen in Figure 7.19) which is not ideal for comparing algorithms nor for the reproducibility of this work.
7.12. Further work

- It is thought that a simpler cluster analysis method could be implemented without losing the benefits of the k-means-mark method.

- One might consider a multiple-sample two-dimensional Kolmogorov-Smirnov Test, taking as a starting point the code implementation of Peacock [246].

- Work out how much of the optimality function is noise.

  - The function plotted in Figure 7.19 is currently rather noisy. A limited amount of this will be the result of the random nature of the k-means analysis but even if this was replaced with some method that did not suffer in this way, I suspect that this function would still be rather jagged due to specific artefacts of any chosen dataset. It would be beneficial to consider whether the general form of the structure was robust to different observers, daylight data, and surface data.

- Compare and combine with other colour constancy algorithms.

  - There are a large number of highly effective colour constancy algorithms being developed within the world of computational colour constancy (See Gijsenij et al. [127] for a review). It would be worthwhile reviewing these with a mind to see whether there are any which would be particularly compatible for combination with a melanopsin-based algorithm.

- Reconsider first-level signals.

  - As discussed above, the approach of Maloney and Wandell [211] may prove valuable to this problem.

  - Recent work has also highlighted the value of a luminance signal to colour constancy [39]. It is unclear whether this is anything further than the loose association between luminance and chromaticity noted early in this chapter.

---

28 Code: https://uk.mathworks.com/matlabcentral/fileexchange/38617-kstest_2s_2d-x1-x2-alpha
• Consider whether a melanopic transform could explain the unexpected results (melanopsin activation induced LM modulation but not S) of Cao et al. [38].

• Consider that if we do use this signal in natural environments, does the degradation of this signal in artificial environments impact our preference ratings? Could evidence of this mechanism show up in lighting preference indices used in the lighting engineering world?
Chapter 8

General Conclusions

8.1 Summary of Conclusions

1. Assuming the applicability of general damage functions, potential damage may be reduced by using light sources with lower CCTs (Figure 2.30).

2. Museum professionals currently do not employ this technique, in part due to conflicting results as to how CCT interacts with observer preference (Section 2.3.4).

3. One potential cause for the conflicting results of experiments looking to find a canonical preferred CCT might be if ipRGCs were involved in colour constancy or chromatic adaptation. This possibility was explored with two psychophysical experiments. The results showed no evidence for a strong or simple interaction (Chapters 4 and 5).

4. To address limitations in the performed experiments, and assist with designing future experiments, a computational study was performed to explore whether a melanopsin interaction would be beneficial for colour constancy (Chapter 7). It was found that there was value in a melanopic input to chromatic adaptation or colour constancy. It was also found that the spectral sensitivity of melanopsin is close to optimal for providing this type of signal.

5. In response to this, and observations from psychophysical work, future work is proposed (Section 8.4).
6. A novel method for performing colour constancy experiments with a tablet computer has been provided, and recommendations for further development are suggested (Chapter 6). This method seems valuable since it allows for colour constancy experiments to be performed in natural environments. It was found that overall this methodology is effective, with certain caveats and controls required.

8.2 Contributions to Museum Lighting

The ideal approach would be to find the damage function for the specific material which is of concern, and the SPD of the various illuminants under consideration, and compute a range of bespoke damage index values for those specific illuminants and that specific material. This could be achieved using the code provided at https://github.com/da5nsy/DamageIndex.

However, where this seems like an unreasonable or unpractical undertaking, based on computations performed here (and relying on the applicability of general damage functions) it seems that it would be a wise choice to pick museum lighting of a lower CCT when faced with a choice between two otherwise similar light sources.

It is likely that some CCTs will be preferred over others. However, it seems likely that luminance and colour rendering are more important factors to visual experience. With this in mind, using the choice of CCT as a means to limit damage seems wise.

8.3 Contributions to Vision Science

The psychophysical experiments reported here set out to answer the question ‘Are ipRGCs involved in colour constancy?’ These experiments do not provide evidence of a melanopic input to colour constancy. However, it is noted that these experiments explore only a small fraction of the stimulus space that ipRGCs may act over. It is further noted that these experiments, and other experiments done in this field elsewhere, do not necessarily correspond to the conditions under which one may expect to see an effect. The computational study reported in Chapter 7
aimed to better define what these conditions are most likely to be.

The computational study makes the prediction that colour constancy using a melanopsin-based mechanism should be effective even when there is only one surface present in the scene. Other candidate mechanisms do not share this prediction, and it seems as though this should be a testable hypothesis.

However, since based on the results of Kraft and Brainard [176] it seems likely that colour constancy is achieved through the interplay of a great many factors, and there is not a single panacea to the problem of illuminant variance, special care should be taken. It is likely that in some situations certain mechanisms or cues are given weight, where in another situation, others are prioritised.

Since at this time there is no clear indication of the level at which the visual system might use a melanopic signal, it seems wise to focus future experiments on high-level and naturalistic behavioural tasks such as colour naming and differentiation of surface from illumination changes, rather than techniques which potentially primarily probe low level mechanisms, such as achromatic matching.

8.4 Future work

Future work relevant to the computational study and the tablet method is listed at the end of the respective chapters. A number of recommendations for future sphere-type psychophysical investigations are provided below.

- The experiments reported here relied upon the assumption of cross-retinal adaptation. Whilst this is an interesting area to explore, and certainly worthy of further investigation, a simpler place to start would be to use corresponding locations for adaptation and testing. For example, a screen or other illumination source comprised of a number of concentric circles of different diameters, where the centre and each annulus could be uniquely addressed, would allow an investigator to adapt a particular eccentricity of the retina, and test on that same part of the retina (or a different area). This would have the additional benefit of allowing for observer metameric matches which satisfied the various bands of different spectral sensitivity as eccentricity.
Chapter 8. General Conclusions

• A valuable concept introduced by Spitschan et al. [281] is that of ‘splatter’. This is “the expected amount of contrast on nominally silenced photoreceptor classes for a given modulation around a given background” and is introduced in recognition of the limitations of stimulus control. Including a control condition, or multiple control conditions, where a conservative estimate of the amount of cone/rod contrast which is likely to have avoided silencing is presented, can allow for an effect due to the target variable to be distinguished from noise resulting from the limits of stimulus control.

• It is recommended to avoid manual input from observers wherever possible, as this seems to be a considerable source of error. Allen et al. [4] introduced a method for performing perceptual nulling in this type of experiment using an alternative forced choice paradigm, where potential metameric pairs were presented as spatial sinusoidal gratings of one of 4 orientations. An observer had to report the orientation, and the orientations where performance fell to chance levels were selected as ‘functionally metameric’ pairs for each observer. An analogous method to replace the task of achromatic setting might be the method of Smithson and Zaidi [275], whereby observers make category judgements rather than tuning the appearance of a stimulus. This has the advantage that stimuli can be shown for a defined amount of time, avoiding the problem whereby the test stimulus affects the observer’s state of adaptation. Informal discussion with these investigators has suggested that whilst this method garners data with low levels of noise, it is relatively time-consuming.

• It does not appear as though this type of experiment is susceptible to the issues identified by Spitschan et al. [281] whereby cones in the shadow of retinal blood vessels have a different effective spectral sensitivity, and thereby break designed metamerism. This seems only to affect high temporal frequency presentations, which suggests that the perceptual nulling stage may be at
risk. Zele et al. [329] propose a method to avoid this issue by desensitising cones with temporal white noise which is nominally melanopsin silent.

- The experiments reported here were of relatively low luminance, and personal communications with other researchers in the field have suggested that melanopsin is more likely to be active only at relatively high luminances.

- In experiments which explore long term adaptation to multiple nominally metameric conditions it would be beneficial to test throughout the experiment whether metamerism was holding. This could be achieved, for example, with brief repeats of the perceptual nulling task, interleaved into the main task.

- The computational study performed in Chapter 7 identified that an absolute melanopic signal (‘first-level’, to use the terminology used in that chapter) may be an inferior target signal compared to a signal which was in some way normalised. The specific normalisation identified in the computational study was a normalisation by luminance, following the template for computing MB chromaticity values, though further research is advised before settling on this as a new target variable (it is quite plausible that normalisation to another signal might provide a superior signal).

- A final note, which is in recognition of a change to standard practice which has occurred during recent years: future experiments of this type (so long as there is a clear model from which a definite prediction can be made) should pre-register the hypothesis and methodology, since it seems likely that this type of experiment (with many uncontrolled / uncontrollable factors and high levels of noise) may be particularly susceptible to post-hoc analysis.
Appendix A

Colophon

This document has been written with \LaTeX, within Overleaf (www.overleaf.com), based on a template created by Ian Kirker\(^1\).

During writing git was used as a version control software, pushing to GitHub, at https://github.com/da5nsy/Thesis. Upon final submission this repo will be made public, and future students are encouraged to re-use any formatting elements which have been developed on top of the template listed above.

The font used is Linux Libertine, which is a free and open font, available at http://libertine-fonts.org/ or through the \LaTeX\,package ‘libertine’.

Zotero was used for reference management.

I am deeply grateful to the open source projects listed above.

\(^1\)https://github.com/UCL/ucl-latex-thesis-templates
Appendix B

Ethics Application 9357/003
UCL Research Ethics Committee

Note to Applicants: It is important for you to include all relevant information about your research in this application form as your ethical approval will be based on this form. Therefore anything not included will not be part of any ethical approval.

You are advised to read the Guidance for Applicants when completing this form.

<table>
<thead>
<tr>
<th>Application For Ethical Review: Low Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are you applying for an urgent accelerated review?</td>
</tr>
<tr>
<td>If yes, please state your reasons below. Note: Accelerated reviews are for exceptional circumstances only and need to be justified in detail.</td>
</tr>
</tbody>
</table>

| Is this application for a continuation of a research project that already has ethical approval? | Yes ☐ No ☒ |
| For example, a preliminary/pilot study has been completed and is this an application for a follow-up project? |

| If yes, provide brief details (see guidelines) including the title and ethics id number for the previous study: | N/A |

Section A: Application details

<table>
<thead>
<tr>
<th></th>
<th>Title of Project</th>
<th>Chromatic Adaptation Experiments in Spheres</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Proposed data collection start date</td>
<td>20/11/2017</td>
</tr>
<tr>
<td>2</td>
<td>Proposed data collection end date</td>
<td>02/12/2017</td>
</tr>
<tr>
<td>3</td>
<td>Project Ethics Identification Number</td>
<td>9357/003</td>
</tr>
<tr>
<td>4</td>
<td>Principal Investigator</td>
<td>Daniel Garside</td>
</tr>
<tr>
<td>5</td>
<td>Position held (Staff/Student)</td>
<td>PhD Student</td>
</tr>
<tr>
<td>6</td>
<td>Faculty/Department</td>
<td>CEGE</td>
</tr>
<tr>
<td>7</td>
<td>Course Title (if student)</td>
<td>Research PhD</td>
</tr>
<tr>
<td>8</td>
<td>Contact Details</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Email:</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Provide details of other Co-Investigators/Partners/Collaborators who will work on the project.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Note: This includes those with access to the data such as transcribers.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Name: Kees Teunissen</td>
<td>Name: Lindsay MacDonald</td>
</tr>
<tr>
<td></td>
<td>Position held: Honorary Professor</td>
<td></td>
</tr>
</tbody>
</table>
Position held: Senior Research Scientist (Industry Supervisor)  
Faculty/Department: N/A  
Location (UCL/overseas/other UK institution): Philips Lighting Research, Eindhoven, Netherlands  
Email: [redacted]

If you do not know the names of all collaborators, please write their roles in the research.

11 If the project is funded (this includes non-monetary awards such as laboratory facilities)

<table>
<thead>
<tr>
<th>Name of Funder</th>
<th>EPSRC and Philips Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the funding confirmed?</td>
<td>Yes – part of funded PhD</td>
</tr>
</tbody>
</table>

12 Name of Sponsor

The Sponsor is the organisation taking responsibility for the project, which will usually be UCL. If the Sponsor is not UCL, please state the name of the sponsor. N/A

13 If this is a student project

<table>
<thead>
<tr>
<th>Supervisor Name</th>
<th>Stuart Robson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position held</td>
<td>Head of Dept.</td>
</tr>
<tr>
<td>Faculty/Department</td>
<td>CEGE</td>
</tr>
<tr>
<td>Contact details</td>
<td>[redacted]</td>
</tr>
</tbody>
</table>

Section B: Project details

The following questions relate to the objectives, methods, methodology and location of the study. Please ensure that you answer each question in lay language.

14 Provide a brief (300 words max) background to the project, including its intended aims.

The aim of the PhD project is to improve our understanding of chromatic adaptation (the process whereby the sensitivity of human visual system changes in response to changes in the colour of ambient lighting), specifically with the aim of improving museum lighting, where extreme colours and low overall levels are required for conservation reasons.

Chromatic adaptation plays a role in providing stable perceptions of objects (objects don’t change drastically in appearance when lighting changes), which is important for object understanding and recognition. It is understood that this ability is provided by a scaling, or calibration, of retinal cone cell outputs (the cells which provide our visual sensation at most luminances).

The incumbent theory states that retinal cone cells are self-calibrating due to bleaching (reaching a state of saturation). Our hypothesis is that another class of retinal cells, ‘intrinsically photosensitive retinal ganglion cells’ (ipRGCs), play a role by calibrating signals from the other retinal cells before they are passed on to the optic nerve. These cells seem to have the appropriate attributes to calibrate for slow changes in the colour of daylight.
Through a carefully designed experiment, where participants observe a set-up under different lighting conditions, designed to differentially affect the ipRGCs, and report their observations, we shall be able to gain an insight as to whether this is indeed the case. If we see performance which corresponds that which we would expect to see if ipRGCs were involved in the process, this will provide evidence in support of this theory.

If this proves to be the case, it would vastly add to our understanding of the human visual system, and have widespread implications in the related disciplines of lighting engineering and colour reproduction.

15 **Methodology & Methods** (tick all that apply)

|☐| Interviews*
|☐| Focus groups*
|☐| Questionnaires (including oral questions)*
|☐| Action Research
|☐| Observation
|☐| Documentary analysis (including use of personal records)
|☐| Audio/visual recordings (including photographs)

*Attach copies to application (see below).

|☐| Collection/use of sensor or locational data
|☒| Controlled Trial
|☐| Intervention study (including changing environments)
|☐| Systematic review
|☐| Secondary data analysis – *(See Section D)*
|☐| Advisory/consultation groups
|☒| Other, give details:

The proposed experiment falls into the category of 'psychophysical study, a common method is vision science. Definition: "Psychophysics quantitatively investigates the relationship between physical stimuli and the sensations and perceptions they produce."

16a **Provide – in lay person’s language - an overview of the project;** focusing on your methodology and including information on what data/samples will be taken (including a description of the topics/questions to be asked), how data collection will occur and what (if relevant) participants will be asked to do. This should include a justification for the methods chosen. *(500 words max)*

This experiment aims to provide information about the eye and the brain through the presentation of visual stimuli on a computer screen, where the participant makes observations and reports these observations by moving two handheld sliders.

The experiment is a psychophysical study (psychophysics being the branch of psychology which deals with the relationship between physical stimuli and the sensations and perceptions they produce) of a type commonly used in studies of this type; that of ‘achromatic setting’. In an experiment of this kind, an observer views a scene (referred to as an adapting stimulus) for an extended period (generally a broad flat field of colour provided by a computer screen or some sort of lighting booth) before setting the appearance of a second stimulus such that it appears achromatic (without colour: black, grey or white).

Using multiple different adapting stimuli, the relative effect of each adapting stimulus can be determined. It would be expected that the effect of an adapting stimulus would be to shift the observers perceived neutral point towards the colour of the adapting stimulus. For example, if an observer viewed a strongly blue adapting stimulus for a length of time, and then immediately modified the appearance of a second stimulus to appear...
grey to them, their chosen colours would normally be objectively blue-er than they would be under normal circumstances.

In this case, the observer shall look into an illuminated sphere of roughly 70cm diameter, where the attributes of the adapting stimulus (the interior walls of the sphere) shall be determined by the lighting (provided through a hole on the top of the sphere), whilst using handheld sliders to control the colour of an LCD screen visible through a small hole on the far side of the sphere.

The observer's task will be to modify the colour of the visible part of the screen until it appears achromatic. Upon pressing a button to confirm their satisfaction with the achromacy of the colour, a new random colour will be presented on the screen, and the observer will again be asked to adjust this new colour until it appears achromatic. This process will repeat 150 times per session (which should take observers between 30 minutes and 1 hour). Continuing the experiment for this length of time serves multiple purposes. Firstly, if there is a slow ramp up to a stable adaptation point (as is the case with rods which take roughly 40 minutes to fully adapt) then this shall be captured in the data. Secondly, if adaptation continued beyond an hour (as some have hypothesised it may) then an impression of this shall be captured by looking for an ongoing trend line across the time of the experiment. Finally, data collecting using a method such as this is generally very noisy. Taking 150 recordings for each session shall allow us to calculate reasonably accurate averages, and also to explore the influence of other factors on the introduction of this noise.

There will exist two distinct experimental conditions. Whilst both should appear identical to observers, they will provide differing levels of ipRGC activation. This type of design is referred to as a 'silent substitution' design, and is possible through the use of narrowband LEDs in combination for the illumination. Observers will repeat observations of the two conditions. Sessions will be held on successive days. Each observer will therefore be asked to complete 4 sessions of up to one hour in length, one per day. Additionally, they will be asked to attend an initial set-up session in order to set the illumination conditions such that they appear identical for each individual observer.

A result which showed little distinction between the two conditions would provide evidence for the incumbent theory (that the cone cells are self-regulating), whereas a result that showed great distinction for these two visually identical conditions would provide evidence for the theory that non-cone cells, such as ipRGCs, are involved in the process of chromatic adaptation.

16b Attachments
If applicable, please attach a copy of any interview questions/workshop topic guides/questionnaires/test (such as psychometric), etc and state whether they are in final or draft form. N/A

17 Please state which code of ethics (see Guidelines) will be adhered to for this research (for example, BERA, BPS, etc).
BPS

Location of Research
18 Please indicate where this research is taking place.
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 19 | If the research includes work outside the UK, is ethical approval in the host country (local ethical approval) required? (See Guidelines.)  
|   | Yes ☐ No ☐  
|   | If no, please explain why local ethical approval is not necessary.  
|   | If yes, provide details below including whether the ethical approval has been received.  
|   | Note: Full UCL ethical approval will not be granted until local ethical approval (if required) has been evidenced. |
| 20 | If you (or any members of your research team) are travelling overseas in person are there any concerns based on governmental travel advice ([www.fco.gov.uk](http://www.fco.gov.uk)) for the region of travel?  
|   | Yes ☐ No ☐  
|   | Note: Check [www.fco.gov.uk](http://www.fco.gov.uk) and submit a travel insurance form to UCL Finance (see application guidelines for more details). This can be accessed here: [https://www.ucl.ac.uk/finance/secure/fin_acc/insurance.htm](https://www.ucl.ac.uk/finance/secure/fin_acc/insurance.htm) (You will need your UCL login details.) |
| 21 | State the location(s) where the research will be conducted and data collected. For example public spaces, schools, private company, using online methods, postal mail or telephone communications.  
|   | UCL, Chadwick Building, Room B06 |
| 22 | Does the research location require any additional permissions (e.g. obtaining access to schools, hospitals, private property, non-disclosure agreements, access to biodiversity permits (CBD), etc.)?  
|   | Yes ☐ No ☒  
|   | If yes, please state the permissions required. |
| 23 | Have the above approvals been obtained?  
|   | Yes ☐ No ☐  
|   | If yes, please attach a copy of the approval correspondence.  
|   | If not, confirm they will be obtained prior to data collection. Yes ☐ No ☐ |

**Section C: Details of Participants**

In this form ‘participants’ means human participants and their data (including sensor/locational data, observational notes/images, tissue and blood samples, as well as DNA).

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 24 | Does the project involve the recruitment of participants?  
|   | Yes ☒ Complete all parts of this Section.  
|   | No ☐ Move to Section D. |
### Participant Details

25 Approximate maximum number of participants required: 10

| Approximate upper age limit: 30 | Lower age limit: 18 |

**Justification for the age range and sample size:**

Age range: Above the age of 30 the spectral sensitivity of vision changes as a result of yellowing of the lens and other factors. It is a limit of the experimental design of silent substitution experiments such as this that sensitivity to wavelengths approaching 400nm is required, which is increasingly lacking as age increases. The use of participants from a narrow range of ages not only avoids this lack of sensitivity, but also allows for a reduced set-up time since the sensitivity of all observers will be similar.

Sample size: Standard deviation in previous trials was roughly 6 deltaE units, and the expected difference between treatment effects is expected to be roughly 10-15 deltaE units. With 3 observers, the requisite deltaE difference for a 95% confidence interval would be 11.4. With 5, it would be 8.9 and with 10 it would be 6.3. Following this logic, where the resulting deltaE values are only roughly known, using 10 observers seems a fair compromise between confidence and inconvenience.

### Recruitment/Sampling

26 Describe how potential participants will be recruited into the study.

Considering that each observer will be required for one hour per day on multiple days, it seems efficient to recruit participants who would normally attend UCL on a regular basis for other reasons.

For this reason, participants will be recruited from the group of peers (fellow PhD students), and staff members within the department of CEGE, and personal friends who would have reason to be in the proximity.

Participants will be approached in person or via email with details of the study. Special note will be made that no person should feel under any obligation to take part.

### Informed Consent

27a Describe the process you will use when seeking to obtain consent.

Participants will be asked to sign a consent form, having read the information form. See attached consent form.

27b **Attachments** Please list them below:

- Consent form
- Information sheet
- Risk Assessment

27c If you are not intending to seek consent from participants, clarify why below:

N/A

### How will the results be disseminated (including communication of results with participants)?

Potentially:

- Conference presentation(s)
In addition, anonymised data will be made publicly available through a system such as figshare. Participants will be provided, via email, with copies of the above, where copyright allows.

### Section D: Accessing/Using Pre-collected Data

<table>
<thead>
<tr>
<th>Access to data</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
</tr>
<tr>
<td>N/A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accessing pre-collected data</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name of dataset(s):</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Owner of dataset(s) (if applicable):</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Is the data in the public domain?</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
</tr>
<tr>
<td>If not, do you have the owner’s permission/license?</td>
</tr>
<tr>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Is the data anonymised?</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
</tr>
<tr>
<td>If not:</td>
</tr>
<tr>
<td>i. Do you plan to anonymise the data?</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>ii. Do you plan to use individual level data?</td>
</tr>
<tr>
<td>Yes*</td>
</tr>
<tr>
<td>iii. Will you be linking data to individuals?</td>
</tr>
<tr>
<td>Yes*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Is the data sensitive (DPA 1998 definition)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Will you be conducting analysis within the remit it was originally collected for?</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>If not, was consent gained from participants for subsequent/future analysis?</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
</tr>
</tbody>
</table>
### Ethical Issues

37 Please address clearly any ethical issues that may arise in the course of this research and how they will be addressed. Further information and advice can be found in the guidelines.

Participation (power relations) - the use of peers and friends may open the researchers to a situation where participants have a lesser ability to make decisions based on personal preference, such as where a participant may feel compelled to continue with an experiment where they might not wish to. This will be addressed by clearly stating (multiple times) that participants are free to withdraw at any point and that there is no disincentive, of any type, to doing so.

Withholding information (Mild deception) - The observers will not be informed that they are viewing different lighting conditions (reminder - each lighting condition will appear identical). This information will be withheld to avoid bias based on this information. Following the experiment this information will be revealed.

### Risks & Benefits

38 Please state any benefits to participants in taking part in the study (this includes feedback, access to services or incentives).

- Financial reward in the form of Amazon vouchers.

39 Do you intend to offer incentives or compensation, including access to free services)?

- Yes ☒
- No ☐

If yes, specify the amount to be paid and/or service to be offered as well as a justification for this.

- £10 per session, in Amazon vouchers. 5 sessions (4 viewing sessions + initial set-up) = £50.
- If an observer chooses to withdraw partway through, they will be paid for each session completed or part-completed.
- Observers will be required to complete 5 sessions of up to 1 hour in length, at the same time on separate (ideally successive) days.

Whilst the experience of participating in similar experiments has previously been described as ‘meditative’, and ‘quite pleasant’, it is thought that simply the pleasure of taking part, or the goodwill of contributing to science, may not be enough to recruit participants in a timely fashion.

The level of reimbursement considered above is proposed as fair and appropriate considering the level of burden, with thought given to the additional burden of travel to the experimental location of 5 separate days, at specific times of day (each observer will need to do each session at the same time of day throughout their run of 5 sessions, and so this is likely to have an impact on the participants’ ability to schedule the rest of their week).

Providing financial recompense for time seems an appropriate method for recruiting participants in this case, particularly because of the lack of a time constraint in the experiment (participants can rush, or take their time, as they wish). Providing financial recompense should reduce the incentive for participants to rush as they might do had they volunteered their time freely. If a participant rushes a trial, it is likely that their data will be of less value than if they had not.
The mechanism of Amazon vouchers has been chosen to avoid the need to handle cash. Payment will be made in a timely fashion via email following completion of all sessions, or withdrawal from participation.

In summary, whilst ideally observers would freely offer their time and commitment to this experiment, a financial incentive is deemed appropriate considering the burden placed on observers. This will enable the recruitment of participants in a timely fashion, and encourage the participants to deliver a higher quality of data than might be possible otherwise.

40 Please state any risks to participants and how these risks will be managed.
Light levels will be carefully monitored to ensure that they are safe in terms of spectral characteristics and absolute level.

41 Please state any risks to you or your research team and how these risks will be managed.
The researcher will be working in a darkened room, and so any trip hazards will be minimised.

Section F: Data Storage & Security

Please ensure that you answer each question and include all hard and electronic data.

42 Will the research involve the collection and/or use of personal data?
Yes ☒ No ☐

Personal data is data which relates to a living individual who can be identified from that data OR from the data and other information that is either currently held, or will be held by the data controller (the researcher).

This includes:
- any expression of opinion about the individual and any intentions of the data controller or any other person toward the individual.
- sensor, location or visual data which may reveal information that enables the identification of a face, address, etc (some postcodes cover only one property).
- combinations of data which may reveal identifiable data, such as names, email/postal addresses, date of birth, ethnicity, descriptions of health diagnosis or conditions, computer IP address (if relating to a device with a single user).

If you do not have a registration number from Legal Services, please clarify why not:

43 Is the research collecting or using
- sensitive personal data as defined by the UK Data Protection Act (racial or ethnic origin / political opinions / religious beliefs / trade union membership / physical or mental health / sexual life / commission of offences or alleged offences), and/or
- data which might be considered sensitive in some countries, cultures or contexts.

If yes, state whether explicit consent will be sought for its use and what data management measures are in place to adequate manage and protect the data.

No

44 All research projects using personal data must be registered with Legal Services before the data is collected, please provide the Data Protection Registration Number:
### During the project (including the write up and dissemination period)

**45** State what types of data will be generated from this project (i.e. transcripts, videos, photos, audio tapes, field notes, etc).

- Text and matlab files of chosen colour values.
- Digital files with name, age and sex of participants.

**How will data be stored, including where and for how long?** This includes all hard copy and electronic data on laptops, share drives, usb/mobile devices.

This data will be held until the end of the PhD project (~Nov 2018). Additionally, data will be released as open access data on a website such as figshare. Such data will be anonymised.

**Who will have access to the data, including advisory groups and during transcription?**

Initially, access will be limited to the researcher and their supervisors. Once processing is complete, anonymous data will be released as open access data as above.

**46** Do you confirm that all personal data will be stored and processed in compliance with the Data Protection Act 1998 (DPA 1998).

- Yes ☒
- No ☐

If not, please clarify why.

**47** Will personal data be processed or be sent outside of the European Economic Area (EEA)?*

- Yes ☐
- No ☒

If yes, please confirm that there are adequate levels of protection in compliance with the DPA 1998 and state what the arrangements are below.

*Please note that if you store your research data containing identifiable data on UCL systems or equipment (including by using your UCL email account to transfer data), or otherwise carry out work on your research in the UK, the processing will take place within the EEA and will be captured by Data Protection legislation.

### After the project

**48** What data will be stored and how will you keep it secure?

Name, age and sex of participants, and achromatic co-ordinates.

**Where will the data be stored and who will have access?**

The data will be stored on a password protected laptop, in a locked cupboard, behind a code-locked door, behind a key-card access door.

Where appropriate this data will be transferred to others working on the project.

**Will the data be securely deleted?**

- Yes ☒
- No ☐

If yes, please state when this will occur: At the end of the PhD project (~Nov 2018)
Section G: Declaration

I confirm that the information in this form is accurate to the best of my knowledge.

Signature

Date 02/11/2017

If student:
I have met with and advised the student on the ethical aspects of this project design.

Supervisor Name: Stuart Robson

Supervisor Signature

Date: 02/11/2017

Signature of Head of Department (or Chair of the Departmental Ethics Committee)

Part A
I have read the ‘criteria of minimal risk’ as defined on page 3 of the Guidelines (http://ethics.grad.ucl.ac.uk/forms/guidelines.pdf) and I recommend that this application be considered by the Chair of the UCL REC.

Yes ☒ No ☐

Part B
I have discussed this project with the principal researcher who is suitably qualified to carry out this research and I approve it. I am satisfied that** (highlight as appropriate):

1. Data Protection registration:
   ▪ has been satisfactorily completed

2. A risk assessment:
   ▪ has been satisfactorily completed

Will the data be archived for use by other researchers? Yes ☒ No ☐

If yes, please provide further details including whether researchers outside the European Economic Area will be given access.

Anonymised data will be made openly available using an online repository such as figshare. Access will not be geographically limited.
3. Appropriate insurance arrangements are in place and appropriate sponsorship [funding] has been approved and is in place to complete the study.

Yes ☒ No ☐

4. A Disclosure and Barring Service check(s):
   - ☐ has been satisfactorily completed
   - ☐ has been initiated
   - ☐ is not required

Note: Links to details of UCL’s policies on the above can be found at: http://ethics.grad.ucl.ac.uk/procedures.php

**If any of the above checks are not required please clarify why below.

<table>
<thead>
<tr>
<th>Name:</th>
<th>Nicola Christie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature:</td>
<td>[Redacted]</td>
</tr>
<tr>
<td>Date:</td>
<td>14.11.2017</td>
</tr>
</tbody>
</table>

Updated 19.10.2017
Information Sheet

You will be given a copy of this information sheet.

**Title of Project:** Chromatic Adaptation Experiments in Spheres

This study has been approved by the UCL Research Ethics Committee (Project ID Number): 9357/002

**Name**
Daniel Garside

**Work Address**
Chadwick Building, UCL

**Contact Details**

We would like to invite you to participate in this research project.

**Details of Study:**

Whilst in a dark room you will look into an illuminated sphere, and use handheld sliders to vary the colour of a small circle inside the sphere until you are happy that this patch is a neutral colour (black/grey/white). Upon pressing a button to confirm such, a new colour will be presented, and the process shall start anew. This will continue for sessions of one hour.

This research will increase our understanding of how the visual system adapts to different lighting. Participants will be recruited from colleagues and friends within UCL. Participation consists initially of five sessions of less than 1 hour in length.

You are free to withdraw from participation at any point.

Flickering lights will be used in the set-up, if this presents a risk to your health please alert the experimenter immediately. There are no other risks associated with this experiment further than exposure to light of a similar nature to that of a standard computer monitor.

Participants will be provided with any publications resulting from this research, where copyright allows.

Data will be published online in an anonymous form (without name, age, sex). All data will be collected and stored in accordance with the Data Protection Act 1998.

It is up to you to decide whether to take part or not; choosing not to take part will not disadvantage you in any way. If you do decide to take part you are still free to withdraw at any time and without giving a reason.

At the end of the five sessions, you will receive £50 in amazon vouchers. If you withdraw before completing 5 sessions, you will be reimbursed £10 per session completed or part completed.

Please discuss the information above with others if you wish or ask us if there is anything that is not clear or if you would like more information.
Informed Consent Form

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

Title of Project: Chromatic Adaptation Experiments in Spheres

This study has been approved by the UCL Research Ethics Committee (Project ID Number): 9357/002

If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you to decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.

Participant's Statement

I ____________________

• have read the notes written above and the Information Sheet, and understand what the study involves.
• understand that if I decide at any time that I no longer wish to take part in this project, I can notify the researchers involved and withdraw immediately.
• consent to the processing of my personal information for the purposes of this research study.
• understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.
• agree that the research project named above has been explained to me to my satisfaction and I agree to take part in this study.
• agree that my data, after it has been fully anonymised, can be shared with other researchers
• agree to be contacted in the future by UCL researchers who would like to invite me to participate in follow-up studies.

Signed: ____________________  Date: ____________________
If you are intending to use non-ionising radiation, i.e. lasers, microwave, ultra-violet or other type of electromagnetic energy, please provide details of exposure and how any associated risks will be minimised:

Light safety has been assessed through the use of the implementation of 'Ophthalmic instruments - Fundamental requirements and test methods - Part 2: Light hazard protection (BS EN ISO 15004-2:2007)' written by Manuel Spitschan and available at https://github.com/spitschan/SilentSubstitutionToolbox/tree/master/LightSafety, to make computations on data collected with a Photo Research PR650 of the interior of the sphere on with the illumination on full power.

* ISO MPE Analysis
* Light is UNDER ISO 2007 MPE Type 1 continuous limits
* Type 1 continuous corneal irradiance UV weighted (5.4.1.1)
  * Value: 0.000, limit 0.400 (uWatts/cm²)
* Type 1 continuous corneal irradiance UV unweighted (5.4.1.2)
  * Value: 0.016, limit 1000.000 (uWatts/cm²)
* Type 1 continuous aphakic retinal illuminance weighted (5.4.1.3.a)
  * Value: 0.296, limit 220.000 (uWatts/cm²)
* Type 1 continuous aphakic radiance weighted (5.4.1.3.b)
  * Value: 2.223, limit 2000.000 (uWatts/[sr·cm²])
* Type 1 continuous corneal irradiance IR unweighted (5.4.1.3.b)
  * Value: 0.001, limit 20000.000 (uWatts/[sr·cm²])
* Type 1 continuous thermal retinal illuminance weighted (5.4.1.3.a)
  * Value: 0.740, limit 700000.000 (uWatts/cm²)
* Type 1 continuous thermal radiance weighted (5.4.1.3.b)
  * Value: 5.560, limit 5880000.000 (uWatts/[sr·cm²])
* Assumed duration seconds 7200.0, hours 2.0

Declaration

Declaration by Departmental Non-Ionising Radiation Protection Supervisor (DNIRPS) for the department effecting the exposures.

I am satisfied that the type and degree of radiation exposure are appropriate for the research being undertaken, and that appropriate procedures are in place to minimize any associated risk. A written risk assessment has been performed which concludes that risks to critical groups are acceptable.

Name of DNIRPS: Dr Mike Lockyer UCL Laser Protection Officer

Signature: 

Date: 31.10.17
RISK ASSESSMENT FORM
FIELD / LOCATION WORK

The Approved Code of Practice - Management of Fieldwork should be referred to when completing this form
http://www.ucl.ac.uk/estates/safetynet/guidance/fieldwork/acop.pdf

DEPARTMENT/SECTION CEGE

LOCATION(S) B06, CHADWICK BUILDING

PERSONS COVERED BY THE RISK ASSESSMENT DANIEL GARSIDE

BRIEF DESCRIPTION OF FIELDWORK A participant will, in a dark room, look into an illuminated sphere, and use handheld sliders to vary the colour of a small circle inside the sphere until the participant is happy that this patch is a neutral colour (black/grey/white). Upon pressing a button to confirm such, a new colour will be presented, and the process shall start anew. This will continue for sessions of one hour.

Consider, in turn, each hazard (white on black). If NO hazard exists select NO and move to next hazard section. If a hazard does exist select YES and assess the risks that could arise from that hazard in the risk assessment box. Where risks are identified that are not adequately controlled they must be brought to the attention of your Departmental Management who should put temporary control measures in place or stop the work. Detail such risks in the final section.

ENVIRONMENT
The environment always represents a safety hazard. Use space below to identify and assess any risks associated with this hazard
Examples of risk: adverse weather, illness, hypothermia, assault, getting lost.
Is the risk high / medium / low?

Basement room B06
It will be dark, caution will be taken not to trip

CONTROL MEASURES
Indicate which procedures are in place to control the identified risk

People not familiar with the room layout will be seated before lighting is turned off

EMERGENCIES
Where emergencies may arise use space below to identify and assess any risks
Examples of risk: loss of property, loss of life

In case of fire or accident we should be able to quickly exit the environment in the same manner as a normal staff member would.

CONTROL MEASURES
Indicate which procedures are in place to control the identified risk

Participants will be made aware of the fire escape route.
**EQUIPMENT**

<table>
<thead>
<tr>
<th>Is equipment used?</th>
<th>Yes</th>
<th>If ‘No’ move to next hazard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examples of risk: inappropriate, failure, insufficient training to use or repair, injury. Is the risk high / medium / low?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The risks associated with the equipment to be used might be electrical or optical. Electrically, all items are low voltage and should not come anywhere near a participant. Optically, tests have been carried out before experiments take place to confirm that light levels are below the limits specified in 'Ophthalmic instruments - Fundamental requirements and test methods - Part 2: Light hazard protection (BS EN ISO 15004-2:2007)'.

**CONTROL MEASURES**

- Indicate which procedures are in place to control the identified risk

| | the departmental written Arrangement for equipment is followed |
| | participants have been provided with any necessary equipment appropriate for the work |
| | all equipment has been inspected, before issue, by a competent person |
| | all users have been advised of correct use |
| | special equipment is only issued to persons trained in its use by a competent person |

**OTHER CONTROL MEASURES:** please specify any other control measures you have implemented:

---

**LONE WORKING**

<table>
<thead>
<tr>
<th>Is lone working a possibility?</th>
<th>No</th>
<th>If ‘No’ move to next hazard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examples of risk: difficult to summon help. Is the risk high / medium / low?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Low

**CONTROL MEASURES**

- Indicate which procedures are in place to control the identified risk

| | the departmental written Arrangement for lone/out of hours working for field work is followed |
| | lone or isolated working is not allowed |
| | location, route and expected time of return of lone workers is logged daily before work commences |
| | all workers have the means of raising an alarm in the event of an emergency, e.g. phone, flare, whistle |
| | all workers are fully familiar with emergency procedures |

**OTHER CONTROL MEASURES:** please specify any other control measures you have implemented:
There will be 2 people present or in close contact during all sessions.
**ILL HEALTH**

The possibility of ill health always represents a safety hazard. Use space below to identify and assess any risks associated with this Hazard.

Examples of risk: injury, asthma, allergies. Is the risk high / medium / low?

Low

**CONTROL MEASURES**

Indicate which procedures are in place to control the identified risk

- an appropriate number of trained first-aiders and first aid kits are present on the field trip
- all participants have had the necessary inoculations / carry appropriate prophylactics
- participants have been advised of the physical demands of the trip and are deemed to be physically suited
- participants have been adequate advice on harmful plants, animals and substances they may encounter
- participants who require medication have advised the leader of this and carry sufficient medication for their needs

**OTHER CONTROL MEASURES**: please specify any other control measures you have implemented:

Should ill health occur, experiments can be stopped immediately.

**TRANSPORT**

Will transport be required

- NO  
- YES

Move to next hazard

Use space below to identify and assess any risks

Examples of risk: accidents arising from lack of maintenance, suitability or training

Is the risk high / medium / low?

Above: NO

**CONTROL MEASURES**

Indicate which procedures are in place to control the identified risk

- only public transport will be used
- the vehicle will be hired from a reputable supplier
- transport must be properly maintained in compliance with relevant national regulations
- drivers comply with UCL Policy on Drivers [http://www.ucl.ac.uk/hr/docs/college_drivers.php](http://www.ucl.ac.uk/hr/docs/college_drivers.php)
- drivers have been trained and hold the appropriate licence
- there will be more than one driver to prevent driver/operator fatigue, and there will be adequate rest periods
- sufficient spare parts carried to meet foreseeable emergencies

**OTHER CONTROL MEASURES**: please specify any other control measures you have implemented:

**DEALING WITH THE PUBLIC**

Will people be dealing with public

- No

If ‘No’ move to next hazard

If ‘Yes’ use space below to identify and assess any risks

Examples of risk: personal attack, causing offence, being misinterpreted. Is the risk high / medium / low?

**CONTROL MEASURES**

Indicate which procedures are in place to control the identified risk

- all participants are trained in interviewing techniques
- interviews are contracted out to a third party
- advice and support from local groups has been sought
- participants do not wear clothes that might cause offence or attract unwanted attention
- interviews are conducted at neutral locations or where neither party could be at risk

**OTHER CONTROL MEASURES**: please specify any other control measures you have implemented:

FIELDWORK 3 May 2010
### WORKING ON OR NEAR WATER

<table>
<thead>
<tr>
<th>Will people work on or near water?</th>
<th>No</th>
<th>If ‘No’ move to next hazard</th>
</tr>
</thead>
</table>

Examples of risk: drowning, malaria, hepatitis A, parasites. Is the risk high / medium / low?

#### CONTROL MEASURES

- lone working on or near water will not be allowed
- coastguard information is understood; all work takes place outside those times when tides could prove a threat
- all participants are competent swimmers
- participants always wear adequate protective equipment, e.g. buoyancy aids, wellingtons
- boat is operated by a competent person
- all boats are equipped with an alternative means of propulsion e.g. oars
- participants have received any appropriate inoculations

OTHER CONTROL MEASURES: please specify any other control measures you have implemented:

### MANUAL HANDLING (MH)

<table>
<thead>
<tr>
<th>Do MH activities take place?</th>
<th>No</th>
<th>If ‘No’ move to next hazard</th>
</tr>
</thead>
</table>

Examples of risk: strain, cuts, broken bones. Is the risk high / medium / low?

#### CONTROL MEASURES

- the departmental written Arrangement for MH is followed
- the supervisor has attended a MH risk assessment course
- all tasks are within reasonable limits, persons physically unsuited to the MH task are prohibited from such activities
- all persons performing MH tasks are adequately trained
- equipment components will be assembled on site
- any MH task outside the competence of staff will be done by contractors

OTHER CONTROL MEASURES: please specify any other control measures you have implemented:

FIELDWORK 4 May 2010
### SUBSTANCES

<table>
<thead>
<tr>
<th>Will participants work with substances</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examples of risk: ill health - poisoning, infection, illness, burns, cuts. Is the risk high / medium / low?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### CONTROL MEASURES

Indicate which procedures are in place to control the identified risks

- [ ] the departmental written Arrangements for dealing with hazardous substances and waste are followed
- [ ] all participants are given information, training and protective equipment for hazardous substances they may encounter
- [ ] participants who have allergies have advised the leader of this and carry sufficient medication for their needs
- [ ] waste is disposed of in a responsible manner
- [ ] suitable containers are provided for hazardous waste

**OTHER CONTROL MEASURES:** please specify any other control measures you have implemented:

### OTHER HAZARDS

Have you identified any other hazards? | No | If ‘No’ move to next section |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hazard:</td>
<td></td>
<td>If ‘Yes’ use space below to identify and assess any risks</td>
</tr>
<tr>
<td>Risk: is the risk</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### CONTROL MEASURES

Give details of control measures in place to control the identified risks

### OTHER HAZARDS

Have you identified any other hazards? | No | If ‘No’ move to next section |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hazard:</td>
<td></td>
<td>If ‘Yes’ use space below to identify and assess any risks</td>
</tr>
<tr>
<td>Risk: is the risk</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### CONTROL MEASURES

Have you identified any risks that are not adequately controlled? | NO | Move to Declaration |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td>Use space below to identify the risk and what action was taken</td>
<td></td>
</tr>
</tbody>
</table>

Is this project subject to the UCL requirements on the ethics of Non-NHS Human Research? | Yes |

If yes, please state your Project ID Number 9357/003

For more information, please refer to: [http://ethics.grad.ucl.ac.uk/](http://ethics.grad.ucl.ac.uk/)

### DECLARATION

The work will be reassessed whenever there is a significant change and at least annually. Those participating in the work have read the assessment.

Select the appropriate statement:

- [ ] I the undersigned have assessed the activity and associated risks and declare that there is no significant residual risk
- [ ] I the undersigned have assessed the activity and associated risks and declare that the risk will be controlled by the method(s) listed above

**NAME OF SUPERVISOR** Stuart Robson

**SIGNATURE OF SUPERVISOR**

**DATE** 02/11/2017

**FIELDWORK** 5 May 2010
Appendix C

Ethics Application 9357/001
APPLICATION FORM

SECTION A APPLICATION DETAILS

<table>
<thead>
<tr>
<th>A1</th>
<th>Project Title: Estimating chromatic adaptation in a museum environment using achromatic point setting with a tablet computer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Date of Submission: 20/06/2016</td>
</tr>
<tr>
<td></td>
<td>Proposed Start Date: 29/06/16</td>
</tr>
<tr>
<td></td>
<td>UCL Ethics Project ID Number: 9357/001</td>
</tr>
<tr>
<td></td>
<td>Proposed End Date: 02/07/16</td>
</tr>
</tbody>
</table>

If this is an application for classroom research as distinct from independent study courses, please provide the following additional details:

<table>
<thead>
<tr>
<th>Course Title:</th>
<th>Course Number:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A2</th>
<th>Principal Researcher</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full Name: Stuart Robson</td>
</tr>
<tr>
<td></td>
<td>Position Held: Head of Dept.</td>
</tr>
<tr>
<td></td>
<td>Email:</td>
</tr>
<tr>
<td></td>
<td>Telephone:</td>
</tr>
<tr>
<td></td>
<td>Fax: N/A</td>
</tr>
</tbody>
</table>

Declaration To be Signed by the Principal Researcher

- I have met with and advised the student on the ethical aspects of this project design (applicable only if the Principal Researcher is not also the Applicant).
- I understand that it is a UCL requirement for both students & staff researchers to undergo Disclosure and Barring Service (DBS) Checks when working in controlled or regulated activity with children, young people or vulnerable adults. The required DBS Check Disclosure Number(s) is: N/A
- I have obtained approval from the UCL Data Protection Officer stating that the research project is compliant with the Data Protection Act 1998. My Data Protection Registration Number is: N/A (Anonymised data)
- I am satisfied that the research complies with current professional, departmental and university guidelines including UCL’s Risk Assessment Procedures and insurance arrangements.
- I undertake to complete and submit the ‘Continuing Review Approval Form’ on an annual basis to the UCL Research Ethics Committee.
- I will ensure that changes in approved research protocols are reported promptly and are not initiated without approval by the UCL Research Ethics Committee, except when necessary to eliminate apparent immediate hazards to the participant.
- I will ensure that all adverse or unforeseen problems arising from the research project are reported in a timely fashion to the UCL Research Ethics Committee.
- I will undertake to provide notification when the study is complete and if it fails to start or is abandoned.
Applicant(s) Details (if Applicant is not the Principal Researcher e.g. student details):

Full Name: Daniel Garside
Position Held: MPhil/PhD Candidate

Email: [Redacted]
Telephone: [Redacted]
Fax: N/A

Full Name:
Position Held:
Address:
Email: [Redacted]
Telephone: [Redacted]
Fax: [Redacted]

Sponsor/ Other Organisations Involved and Funding

a) Sponsor: [ ] UCL [ ] Other institution
   If your project is sponsored by an institution other than UCL please provide details: EPSRC/Philips iCASE supported by British Museum

b) Other Organisations: If your study involves another organisation, please provide details. Evidence that the relevant authority has given permission should be attached or confirmation provided that this will be available upon request. Grant Museum and British Museum.

c) Funding: What are the sources of funding for this study and will the study result in financial payment or payment in kind to the department or College? If study is funded solely by UCL this should be stated, the section should not be left blank. Funding as part of doctoral iCASE project, provided by EPSRC/Philips

Signature of Head of Department [or Chair of the Departmental Ethics Committee]
(This must not be the same signature as the Principal Researcher)

A. I have discussed this project with the principal researcher who is suitably qualified to carry out this research and I approve it.

I am satisfied that [please highlight as appropriate]:

(1) Data Protection registration:
   - has been satisfactorily completed
   - has been initiated
   - is not required

(2) a risk assessment:
   - has been satisfactorily completed
   - has been initiated

(3) appropriate insurance arrangements are in place and appropriate sponsorship [funding] has been approved and is in place to complete the study. [ ] Yes [ ] No

(4) a Disclosure and Barring Service check(s):
   - has been satisfactorily completed
B1 Please provide a brief summary of the project in **simple prose** outlining the intended value of the project, giving necessary scientific background (max 500 words).

Colour constancy refers to the stable perception of object colour appearance, in spite of a change in illumination causing a change in the nature of the stimuli reaching an observer’s eye. The process by which we adapt to this change in stimulus is the subject of this investigation. This investigation will trial a new method for investigating this subject, and will also look to further the understanding of how humans adapt in mixed lighting environments. This knowledge should be valuable in the creation of general models of visual appearance, and for architectural lighting design.

**PRINT NAME:** NICOLA CHRISTIE  
**DATE:** 29.06.2016

B2 Briefly characterise in **simple prose** the research protocol, type of procedure and/or research methodology (e.g. observational, survey research, experimental). Give details of any samples or measurements to be taken (max 500 words).

The new proposed method requires an observer to select (with their finger) the most neutral point on a tablet computer screen displaying a range of colours. Each participant will be required to respond to roughly 10 stimuli. It is expected that this will take less than one minute. A subset of observers will also perform previously documented colour constancy tasks upon the same tablet, with comparable level and mode of interaction, and time requirements.

The data collected will consist of response locations on the touch interface, and the corresponding colourimetric co-ordinates. Metadata related to the participants’ demographic will also be collected. The level of metadata has been carefully considered such that only vital metadata is recorded, and that such data will be non-identifying. (To be collected: Age, Gender, Location directly prior to participation, nationality, mother tongue, education level, details of known visual impairments, whether tinted eyeglasses had been worn recently)

*Attach any questionnaires, psychological tests, etc. (a standardised questionnaire does not need to be attached, but please provide the name and details of the questionnaire together with a published reference to its prior usage).*
B3 Where will the study take place (please provide name of institution/department)?
If the study is to be carried out overseas, what steps have been taken to secure research and ethical permission in the study country?
Is the research compliant with Data Protection legislation in the country concerned or is it compliant with the UK Data Protection Act 1998?

The study will initially take place in the public museum space at the Grant Museum. Letter of support from the institution is attached. Following an initial phase at the Grant Museum the experiment will be repeated at the British Museum (the project partner), in various public gallery spaces. The British Museum’s project representative supports the experiment, and arrangements are currently being made for official institutional support and logistical considerations.

The data will be non-identifying. The data will be stored physically and digitally on secure devices within the British Museum and UCL. The anonymised data will later be publicly hosted on the UCL Research Data Storage Service (or a comparable UCL or non-UCL service), such that other investigators might openly access and use the data.

B4 Have collaborating departments whose resources will be needed been informed and agreed to participate?
Attach any relevant correspondence.

All resources required will be available from CEGE.

B5 How will the results be disseminated, including communication of results with research participants?

Broad dissemination: results to be presented at AIC2016 (conference) in Sep 2016.

Communication to research participants: At the end of each run, feedback on input will be provided (there is no ‘right’ answer, but an overview of inputs will be output). If participants wish to be informed of experimental results of the entire group, they will be provided with a business card which has a link with which to sign up to a mailing list where news of publications will be communicated. Signing up to this communication will be very clearly at the volition of the participant.

B6 Please outline any ethical issues that might arise from the proposed study and how they are be addressed. Please note that all research projects have some ethical considerations so do not leave this section blank.

Colour blind observers - it is possible that results from observers with colour anomalous vision (colourblind observers) might be different to those from colour ‘normal’ observers. The approach taken in similar experiments has been to record as metadata whether an observer knows of any vision abnormalities, and then later analyse whether for the specific experiment there is a meaningful effect of such abnormalities. (It has often been the case that there has been negligible functional difference, and so data from these observers has been retained.)

Minors will not be excluded from the experiment, with the exception of those who might be presumed to not understand the experimental procedure. This is the root of the minimum age listed below of 7. Minors will not be approached, not directed in the experimental procedure, nor supervised, nor educated. In the case that a child wishes to take part in the experiment, communication and instruction will be directed to that child's parent or guardian, in order that the experimenter will not be in a position of 'instructor'.

The exclusion of children would be undesirable on two counts. 1- Exclusion (where people actively want to participate, particularly as a family group) is awkward and might be seen as derogatory. 2 - age is a meaningful variable in the experiment.

DBS check is not required as this research does not involve the 'teaching, training or instruction' of minors or vulnerable adults. In the situation that minors or vulnerable adults would be involved in the research, they would participate explicitly under the instruction/supervision of a parent/carer. Specifically in regards to UK law, this research does not constitute a 'Regulated' activity as defined by the 'Safeguarding Vulnerable Groups Act 2006', (2006 c. 47, SCHEDULE 4, Part 1, Activities, Section 2 (1)) (http://www.legislation.gov.uk/ukpga/2006/47/schedule/4/paragraph/2); does not occur in a location listed in the same document in section Schedule 4, Part 1, Establishments, Section 3; and is not performed by a

Data Protection registration not required as all data will be completely non-identifying. UCL Legal Services, Form 2: Research Registration Form: “Projects using anonymised data do not have to be registered with the Data Protection Team and you do not have to worry about compliance with the Act.”

SECTION C DETAILS OF PARTICIPANTS

C1 Participants to be studied

C1a. Number of volunteers: 200

<table>
<thead>
<tr>
<th>Upper age limit:</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower age limit:</td>
<td>7</td>
</tr>
</tbody>
</table>

C1b. Please justify the age range and sample size:
The sample size listed above is a goal size, but a lower number would still be valuable for the experimental exploration.

Historically it is normal for vision experiments to run with <30 observers. Recently however, in part due to the increasing abilities of technology there has been a trend for increasing the number of participants in order to increase the statistical significance of results.

Considering the minimal time and effort required from observers to participate, compared with the amount of time and effort involved in setting up the experiment, and also considering that the effect of time of day/day of the week would be studied, it makes sense to envisage a situation where for several days the experiment was manned, with the number of observers being dependent upon the number of interested visitors who were willing to participate.

C2 If you are using data or information held by a third party, please explain how you will obtain this. You should confirm that the information has been obtained in accordance with the UK Data Protection Act 1998.

N/A

C3 Will the research include children or vulnerable adults such as individuals with a learning disability or cognitive impairment or individuals in a dependent or unequal relationship? ☐ Yes ☐ No

How will you ensure that participants in these groups are competent to give consent to take part in this study? If you have relevant correspondence, please attach it.

This research will not specifically aim to work with the above groups, but there seems to be no reason to exclude the above groups. Where there are uncertainties about an individual’s ability to consent to participation, that individual will not be allowed to participate.
**C4**

**Will payment or any other incentive, such as gift service or free services, be made to any research participant?**

- Yes  ☒
- No 

If yes, please specify the level of payment to be made and/or the source of the funds/gift/free service to be used.

N/A

Please justify the payment/other incentive you intend to offer.

There will be no incentive for taking part in the experiment, other than the opportunity to contribute to a scientific experiment. It is envisaged that no incentive is required considering the educational nature of a museum visit and the minimal amount of effort required for participation on the part of the participant.

---

**C5**

**Recruitment**

(i) Describe how potential participants will be identified:

The participants will be self identifying from the group of ‘British Museum Visitors’. The experimenter (Daniel Garside) will wear clothing which clearly indicates ‘I’m running a scientific experiment’ (probably a UCL T-shirt), and it is expected that members of the public will approach the experimenter, particularly upon seeing other participants participating.

(ii) Describe how potential participants will be approached:

As above, it is likely that they will not be approached, and that in fact they will approach the experimenter. If this does not occur, then gallery visitors might be approached (with as little bias as possible) and presented with a question such as ‘would you like to play a quick game in the name of science?’

(iii) Describe how participants will be recruited:

See above.

Attach recruitment emails/adverts/webpages. A data protection disclaimer should be included in the text of such literature.

---

**C6**

**Will the participants participate on a fully voluntary basis?**  ☒ Yes  ☒ No

**Will UCL students be involved as participants in the research project?**  ☒ Yes  ☒ No

If yes, care must be taken to ensure that they are recruited in such a way that they do not feel any obligation to a teacher or member of staff to participate.

Please state how you will bring to the attention of the participants their right to withdraw from the study without penalty?

See attached verbal experiment introduction.

---

**C7**

**CONSENT**

Please describe the process you will use when seeking and obtaining consent.

The attached consent form will to read aloud to participants before the experiment, and verbal consent will be sought through repetition of phrases. (See final section of ‘Information and Consent’) Printed copies of this information will be available for participants to take after the experiment.

A copy of the participant information sheet and consent form must be attached to this application. For your convenience proformas are provided in C10 below. These should be filled in and modified as necessary.

In cases where it is not proposed to obtain the participants informed consent, please explain why below.
### C8 Will any form of deception be used that raises ethical issues? If so, please explain.

At the start of the experiment simple instructions and guidelines for participation will be verbally explained. The full nature of the experiment as it relates to colour constancy will not be explained before participation, as this would likely bias the results. However, it is considered unlikely that this level of detail would be requested, and where it were it could be provided after the experiment.

### C9 Will you provide a full debriefing at the end of the data collection phase?  
- [ ] Yes  
- [x] No

If ‘No’, please explain why below.

It is thought that the participants will not require debriefing as the experiment is relatively quick and simple to perform, and should have no lasting effect.

### C10 Information Sheets And Consent Forms

A poorly written Information Sheet(s) and Consent Form(s) that lack clarity and simplicity frequently delay ethics approval of research projects. The wording and content of the Information Sheet and Consent Form must be appropriate to the age and educational level of the research participants and clearly state in simple non-technical language what the participant is agreeing to. Use the active voice e.g. “we will book” rather than “bookings will be made”. Refer to participants as “you” and yourself as “I” or “we”. An appropriate translation of the Forms should be provided where the first language of the participants is not English. If you have different participant groups you should provide Information Sheets and Consent Forms as appropriate (e.g. one for children and one for parents/guardians) using the templates below. Where children are of a reading age, a written Information Sheet should be provided. When participants cannot read or the use of forms would be inappropriate, a description of the verbal information to be provided should be given. Please ensure that you trial the forms on an age-appropriate person before you submit your application.

#### Information Sheet for in Research Studies

You will be given a copy of this information sheet.

**Title of Project:**

This study has been approved by the UCL Research Ethics Committee (Project ID Number):

**Name**

**Work Address**

**Contact Details**  
(“For students, we strongly advise against the use of a personal contact number”)
We would like to invite you to participate in this research project.

Details of Study:

Please discuss the information above with others if you wish or ask us if there is anything that is not clear or if you would like more information.

It is up to you to decide whether to take part or not; choosing not to take part will not disadvantage you in any way. If you do decide to take part you are still free to withdraw at any time and without giving a reason.

All data will be collected and stored in accordance with the Data Protection Act 1998.

Thank you for reading this information sheet and for considering take part in this research.

When you have completed your Information Sheet, please DELETE the advice section below from your application form before submitting it to the Committee.

Details of Study MUST include the following:

- Aims of the research and possible benefits.
- Who you are recruiting
- What will happen if the participant agrees to take part (when, where, how long etc)
- Any risks (e.g. need for disclosure of information to a third party, possibility for distress)
- Possible benefits (it is good practice to offer participants a copy of the final report)
- Arrangements for ensuring anonymity and confidentiality (see optional statements below for examples). To ensure compliance with the Data Protection Act participants must be informed of what information will be held about them and who will have access to it (this relates to information that is identifiable or could potentially be linked back to an individual.)

Statements which researchers MIGHT also include as appropriate:

- A decision to withdraw at any time, or decision not to take part, will not affect the standard of care/education you receive.
- If you agree to take part you will be asked whether you are happy to be contacted about participation in future studies. Your participation in this study will not be affected should you choose not to be re-contacted.
- You may withdraw your data from the project at any time up until it is transcribed for use in the final report (insert date).
- Recorded interviews will be transcribed (written up) and the tape will then be wiped clear.
- If you decide to take part you will be given this information sheet to keep and be asked to sign a consent form.
- Submission of a completed questionnaire implies consent to participate.
- As participation is anonymous it will not be possible for us to withdraw your data once you have returned your questionnaire.
- What if I have further questions, or if something goes wrong? If this study has harmed you in any way or if you wish to make a complaint about the conduct of the study you can contact UCL using the details below for further advice and information:
  *Student researchers: Insert the name and full UCL contact address and number of your supervisor.*
  *Staff researchers: Please insert the following: The Chair, *Insert full address details for the UCL Research Ethics Committee, ethics@ucl.ac.uk*

Informed Consent Form for in Research Studies

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

Title of Project:

This study has been approved by the UCL Research Ethics Committee (Project ID Number):

Thank you for your interest in taking part in this research. Before you agree to take part, the person organising the research must explain the project to you.

If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.
Participant's Statement

I

- have read the notes written above and the Information Sheet, and understand what the study involves.
- understand that if I decide at any time that I no longer wish to take part in this project, I can notify the researchers involved and withdraw immediately.
- consent to the processing of my personal information for the purposes of this research study.
- understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.
- agree that the research project named above has been explained to me to my satisfaction and I agree to take part in this study.
- agree that my data, after it has been fully anonymised, can be shared with other researchers [to satisfy Research Council funded projects as Research Councils have changed their guidance regarding data sharing]

Signed: ___________________________  Date: ____________

When you have completed your Informed Consent Form, please DELETE the advice section below from your application form before submitting it to the Committee.

Statements which researchers MIGHT include as appropriate:

- I understand that my participation will be taped/video recorded and I consent to use of this material as part of the project.
- I understand that I must not take part if
- I agree to be contacted in the future by UCL researchers who would like to invite me to participate in follow-up studies.
- I understand that the information I have submitted will be published as a report and I will be sent a copy. Confidentiality and anonymity will be maintained and it will not be possible to identify me from any publications.
- I understand that I am being paid for my assistance in this research and that some of my personal details will be passed to UCL Finance for administration purposes.
- I agree that my non-personal research data may be used by others for future research. I am assured that the confidentiality of my personal data will be upheld through the removal of identifiers.

This is not an exhaustive list and you should consider whether you need to amend any of these statements or design different ones that are more applicable to your research.

SECTION D DETAILS OF RISKS AND BENEFITS TO THE RESEARCHER AND THE RESEARCHED

Have UCL’s Risk Assessment Procedures been followed?  ☒ Yes  ☐ No

If No, please explain.

See attached risk assessment.
Does UCL’s insurer need to be notified about your project before insurance cover can be provided?  □ Yes  □ No

The insurance for all UCL studies is provided by a commercial insurer. For the majority of studies the cover is automatic. However, for a minority of studies, in certain categories, the insurer requires prior notification of the project before cover can be provided.

If Yes, please provide confirmation that the appropriate insurance cover has been agreed. Please attach your UCL insurance registration form and any related correspondence.

This experiment does not fall under the conditions discussed at:
http://ethics.grad.ucl.ac.uk/uclinsurance.php
and therefore insurance is automatic as described above.

Confirmation of this has been provided by Richard Sharp (Departmental Manager for CEGE)

Please state briefly any precautions being taken to protect the health and safety of researchers and others associated with the project (as distinct from the research participants).

I shall stay hydrated and take breaks from standing as required.

Will these participants participate in any activities that may be potentially stressful or harmful in connection with this research?  □ Yes  □ No

If Yes, please describe the nature of the risk or stress and how you will minimise and monitor it.

I cannot envisage a situation whereby this experiment might cause any distress or harm.

Will group or individual interviews/questionnaires raise any topics or issues that might be sensitive, embarrassing or upsetting for participants?

If Yes, please explain how you will deal with this.

No formal interview. After the experiment limited demographic questions will be asked via the tablet computer, with users providing touch response. In the case a participant wishes not to answer a specific question, there will be a clear option for it to be skipped.
**D6** Please describe any expected benefits to the participant.

None.

**D7** Specify whether the following procedures are involved:

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any invasive procedure(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical contact</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any procedure(s) that may cause mental distress</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please state briefly any precautions being taken to protect the health and safety of the research participants.

The screen will be cleaned with antibacterial wipes as required.

**D8** Does the research involve the use of drugs? | Yes | No

If Yes, please name the drug/product and its intended use in the research and then complete Appendix I

N/A

**D9** Will any non-ionising radiation be used on the research participant(s)? | Yes | No

If Yes, please complete Appendix II.

**D10** Are you using a medical device in the UK that is CE-marked and is being used within its product indication? | Yes | No

If Yes, please complete Appendix III.
Please submit either 12 copies (1 original + 11 double sided photocopies) of your completed application form for full committee review or 3 copies (1 original + 2 double sided copies) for chair’s action, together with the appropriate supporting documentation from the list below to the UCL Research Ethics Committee Administrator. You should also submit your application form electronically to the Administrator at: ethics@ucl.ac.uk

<table>
<thead>
<tr>
<th>Documents to be Attached to Application Form (if applicable)</th>
<th>Ticked if attached</th>
<th>Ticked if not relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Section B: Details of the Project</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Questionnaire(s) / Psychological Tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Relevant correspondence relating to involvement of collaborating department/s and agreed participation in the research.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Section C: Details of Participants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Parental/guardian consent form for research involving participants under 18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Participant/s information sheet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Participant/s consent form/s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Advertisement</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Section D: Details of Risks and Benefits to the Researcher and the Researched</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Insurance registration form and related correspondence</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Appendix I: Research Involving the Use of Drugs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Relevant correspondence relating to agreed arrangements for dispensing with the pharmacy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Written confirmation from the manufacturer that the drug/substance has been manufactured to GMP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Proposed volunteer contract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Full declaration of financial or direct interest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Copies of certificates: CTA etc…</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Appendix II: Use of Non-Ionising Radiation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Appendix III: Use Medical Devices</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please note that correspondence regarding the application will normally be sent to the Principal Researcher and copied to other named individuals.
Appendix D

Ethics Application 9357/001 - Amendment
Amending an Approved Application

Should you wish to make an amendment to an approved study, you will need to submit an ‘amendment request’ for the consideration of the Chair of the UCL Research Ethics Committee. Applications can only be amended after ethical approval has been granted.

You will need to apply for an amendment approval if you wish to:

1. Add a new participant group;
2. Add a new research method;
3. Ask for additional data from your existing participants;
4. Remove a group of participants or a research method from the project, and have not yet commenced that part of the project;
5. Apply for an extension to your current ethical approval.

If you need to apply for an amendment approval, please complete the Amendment Approval Request Form on the next page.

When completing the form, please ensure you do the following:

- Clearly explain what the amendment you wish to make is, and the justification for making the change.
- Insert details of any ethical issues raised by the proposed amendments.
- Include all relevant information regarding the change so that the Chair can make an informed decision, and submit a copy of the sections of your application that have changed with all changes highlighted/underlined for clarity.
- You do not need to submit your original application in full again. However, if the changes you wish to make alters several sections of your application form, you are advised to submit this.

Please email a signed electronic copy to the REC Administrator: ethics@ucl.ac.uk

Amendment requests are generally considered within 5-7 days of submission.
# Amendment Approval Request Form

<table>
<thead>
<tr>
<th></th>
<th>Project ID Number: 9357/001</th>
<th>Name and Address of Principal Investigator: Daniel Garside</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Project Title: Estimating chromatic adaptation in a museum environment using achromatic point setting with a tablet computer</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Type of Amendment/s (tick as appropriate)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Research procedure/protocol (including research instruments)</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td>Participant group</td>
<td>☒</td>
</tr>
<tr>
<td></td>
<td>Sponsorship/collaborators</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td>Extension to approval needed (extensions are given for one year)</td>
<td>☒</td>
</tr>
<tr>
<td></td>
<td>Information Sheet/s</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td>Consent form/s</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td>Other recruitment documents</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td>Principal researcher/medical supervisor*</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>☒ - Location</td>
</tr>
<tr>
<td><em>Additions to the research team other than the principal researcher, student supervisor and medical supervisor do not need to be submitted as amendments but a complete list should be available upon request</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 4 | Justification (give the reasons why the amendment/s are needed) |
|   | Running the experiment in an environment where the lighting can be carefully controlled, in order to ask an experimental question which was not possible to test rigorously in the previous environment. The location proposed is PAMELA ([http://www.ucl.ac.uk/transport-institute/lab/PAMELA](http://www.ucl.ac.uk/transport-institute/lab/PAMELA)) since there is a controllable multi-channel LED lighting set-up which is appropriate for this new question. |

| 5 | Details of Amendments (provide full details of each amendment requested, state where the changes have been made and attach all amended and new documentation) |
|   | Location: changing from British Museum to PAMELA. |
|   | Participant group: changing from ‘public’ to ‘colleagues and friends’ (recruited from inside and outside UCL). Whereas the aim was to survey 200 members of the public in the previous application, here we aim to recruit between 5 and 10 participants. |
|   | Dates: Approval was previously given on 11/07/2016 until 11/07/2017. The proposed experimental date at PAMELA is 05/02/2018. (Research procedure: identical procedure, but each participant will complete the experiment multiple times to assess variability. Info sheet will be amended to reflect this.) |

| 6 | Ethical Considerations (insert details of any ethical issues raised by the proposed amendment/s) |
|   | Observers are required to participate for a longer period in this new experiment than any individual member of the public was required to. The experiment is not straining or harmful when undertaken for any reasonable amount of time. For this reason, I consider it appropriate to offer an incentive to participants, to remunerate participants for their time. Participants will be required to attend for 2 hours, and in line with other experiments an amazon gift card of £20 is considered to be appropriate. |

| 7 | Other Information (provide any other information which you believe should be taken into account during ethical review of the proposed changes) |

| Declaration (to be signed by the Principal Researcher) |
| I confirm that the information in this form is accurate to the best of my knowledge and I take full responsibility for it. |
• I consider that it would be reasonable for the proposed amendments to be implemented.
• For student projects, I confirm that my supervisor has approved my proposed modifications.

Signature: 

Date: 29/01/2018

FOR OFFICE USE ONLY:

Amendments to the proposed protocol have been .................. by the Research Ethics Committee.

Signature of the REC Chair:

Date:
References


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES

Two-dimensional-goodness-of-fit-testing-in. (cited: p. 297)


[254] Xudong Qiu, Tida Kumbalasiri, Stephanie M. Carlson, Kwoon Y. Wong, Vanitha Krishna, Ignacio Provencio, and David M. Berson. Induction of photosensitivity by heterologous


REFERENCES


[307] Françoise Viénot, Hans Brettel, Tuong-Vi Dang, and Jean Le Rohellec. Domain of metamers exciting intrinsically photosensitive retinal ganglion cells (ipRGCs) and rods. JOSA A, 29(2):


REFERENCES


REFERENCES


“If, in addition, we feel that all of this has evolved in intimacy with the kinds of order that exist over extended periods of chaos in what we call the outside world, then we can see that there really is no outside world and no inside world. There is just one world. It is, perhaps, a little bit like moss growing on a rock, clinging to it, the tendrils penetrating the crevices in the rock and the cavities of the rock, where the rock/moss combination is the object and not the rock or the moss separately.”

Land [187]